# Measures and Models of Financial Risk

DISSERTATION

zur Erlangung des akademischen Grades
**doctor rerum naturalium**
(dr. rer. nat.)
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
Humboldt-Universität zu Berlin

von
Diplom-Mathematiker Stefan Weber
geboren am 6. Mai 1973 in Hannover

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jürgen Mlynek

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II:
Prof. Dr. Uwe Küchler

Gutachter:

(1) Prof. Dr. Hans Föllmer
(2) Prof. Dr. Frank Riedel
(3) Prof. Dr. Alexander Schied



| | |
|---|---|
| eingereicht am: | 29. Juni 2004 |
| Tag der mündlichen Prüfung: | 12. November 2004 |

## Abstract

In this thesis, we study monetary measures and endogenous models of financial risk. The first part considers two aspects of the quantification of financial risk. We focus on the one hand on the calculation of risk measurements by Monte Carlo simulation. On the other hand, we investigate a particular class of dynamic risk measures. In the second part we analyze two models of financial risk in economies with interacting agents. First, we focus on credit contagion of firms which interact with each other in a network of business partners. Second, we investigate the market interaction of investors with bounded rationality in an evolutionary selection market model.

The simulation of distributions of the value of financial positions is an important issue for financial institutions. If risk measures are evaluated for a simulated distribution instead of the model-implied distribution, the probability of errors of risk measurements need to be analyzed. This topic is investigated in Chapter 1. For distribution-invariant risk measures which are continuous on compacts and for value at risk we derive large deviation bounds. If the approximate risk measurements are based on the empirical distribution of independent samples, the rate function equals the minimal relative entropy under a risk measure constraint. For average value at risk (AVaR) and shortfall risk we solve this minimization problem explicitly.

Chapter 2 provides an axiomatic characterization of dynamic risk measures. We prove a representation theorem and investigate the connection to static risk measures. Two notions of dynamic consistency are proposed. A key insight is that dynamic consistency and the notion of *measure convex sets of probability measures* are intimately related. Measure convexity can be interpreted using the concept of compound lotteries. This leads to a characterization of a class of static risk measures closely connected to shortfall risk.

Chapter 3 investigates credit contagion. Credit contagion refers to the propagation of economic distress from one firm to another. Using methods from the theory of interacting particle systems, we propose a model for these contagion phenomena, assuming they are due to the local interaction of firms in a business partner network. We study aggregate credit losses on large portfolios of financial positions contracted with firms subject to credit contagion. In particular, we provide an explicit Gaussian approximation of the distribution of portfolio losses. We find that contagion processes induce additional fluctuations of losses around their averages, whose size depends on the denseness of the business partner network.

In Chapter 4 we derive a continuous time approximation of the evolutionary market selection model of Blume and Easley (1992). Conditions on the payoff structure of the assets are identified that guarantee convergence. We prove that the continuous time approximation equals the solution of an integral equation in a random environment. For constant asset returns, it reduces to an autonomous ordinary differential equation. We study its long-run behavior using techniques related to Lyapunov functions, and compare our results to the benchmark of profit-maximizing investors.

## Zusammenfassung

Diese Arbeit behandelt die Bemessung und endogene Modellierung von Finanzrisiken. Teil I untersucht neben der Monte Carlo Simulation statischer Risikomaße auch die dynamische Bemessung von Finanzrisiken. In Teil II analysieren wir zwei Modelle mit interagierenden Akteuren. Dabei betrachten wir einerseits Ansteckungsprozesse auf Kreditmärkten, andererseits einen evolutionären Marktselektionsmechanismus.

Für Finanzinstitutionen ist die Simulation von Verteilungen von Finanzpositionen von großer Bedeutung. Werden Risikomaße nicht für die wahre, sondern für die simulierte Verteilung berechnet, ist die Analyse der Wahrscheinlichkeit von Fehlern des so ermittelten Risikos wichtig. Mit dieser Fragestellung befasst sich Kapitel 1. Für verteilungsinvariante Risikomaße, die stetig auf Kompakta sind, und für Value at Risk untersuchen wir große Abweichungen. Beruht die approximative Risikobemessung auf den empirischen Verteilungen unabhängiger Simulationen, ist die Ratenfunktion der großen Abweichungen als eine minimale relative Entropie unter einer Risikomaßnebenbedingung gegeben. Das resultierende Minimierungsproblem lösen wir explizit für Average Value at Risk (AVaR) und Shortfall-Risk.

In Kapitel 2 untersuchen wir dynamische Risikomaße axiomatisch. Wir beweisen einen Darstellungssatz, der den engen Zusammenhang zu statischen Risikomaßen aufzeigt. Wir definieren zwei Arten dynamischer Konsistenz. Es stellt sich heraus, dass eine enge Verbindung zwischen der dynamischen Konsistenz und maßkonvexen Mengen von Wahrscheinlichkeitsmaßen besteht. Maßkonvexität lässt sich auch mit Hilfe von zusammengesetzten Lotterien interpretieren. Dieser Zusammenhang führt auf eine Charakterisierung von statischen Risikomaßen, die eng mit Shortfall-Risk verwandt sind.

Kapitel 3 befasst sich mit Ansteckungsprozessen auf Kreditmärkten. Wir konstruieren mit Hilfe der Theorie interagierender Teilchensysteme ein Modell für Ansteckungsprozesse von lokal interagierenden Geschäftspartnern. Für große Portfolios von Bankkrediten, die an vernetzte und interagierende Firmen vergeben sind, untersuchen wir aggregierte Verluste und leiten eine explizite Gaußsche Approximation der Verlustverteilung her. Die Ansteckungsprozesse induzieren zusätzliche Fluktuation der Verluste, deren Größe vom Grad der Vernetztheit der Ökonomie abhängt.

In Kapitel 4 konstruieren wir eine zeitstetige Approximation eines evolutionären Marktselektionsmodells von Blume and Easley (1992). Wir identifizieren Bedingungen an den Dividendenprozess, die die Konvergenz implizieren. Die zeitstetige Approximation ist Lösung einer Integralgleichung in einer zufälligen Umgebung. Für den Spezialfall zeitlich konstanter Dividendenzahlungen ergibt sich eine autonome gewöhnliche

Differentialgleichung, deren Langzeitverhalten wir mit Hilfe von Lyapunovfunktionen charakterisieren. Unsere Resultate vergleichen wir mit einem Modell von rationalen Investoren.

# Contents

# Introduction

In this dissertation, we study monetary measures and endogenous models of financial risk. The first part of the thesis considers two aspects of the quantification of financial risk. We focus on the one hand on the calculation of risk measurements by Monte Carlo simulation. On the other hand, we investigate a particular class of dynamic risk measures. In the second part we analyze two models of financial risk in economies with interacting agents. First, we focus on credit contagion of firms which interact with each other in a network of business partners. Second, we investigate the market interaction of investors with bounded rationality in an evolutionary selection market model. All chapters of this thesis are self-contained.

### Part I – Measures of Financial Risk

The portfolios of banks consist of financial assets such as stocks, bonds, credits and options. Banks need to manage their risks and are required to respect regulatory constraints. For the quantification of risk associated with these positions banks have to use appropriate measures of risk. While the theory of static risk measures is already well developed, both the implementation of static risk measurements and the dynamic quantification of financial cash flows require additional analysis.

The first chapter of this thesis investigates the calculation of static risk measurements by Monte Carlo methods. Often financial positions are modelled as real-valued random variables. In many cases the associated model distributions are not tractable, but can be simulated by Monte Carlo methods. If then risk measures are evaluated for the simulated instead of the model-implied distributions, the errors of these risk measurements need to be analyzed.

In Chapter 1 we employ the theory of large deviations to study these errors. In a first step, we describe how the error of the risk measurements and large deviations are related. For distribution-invariant risk measures which are continuous on compacts a large deviations principle is an immediate corollary of the contraction principle. Section 1.2.2 analyzes therefore the notion of continuity on compacts. In particular, we prove sufficient conditions in terms of robust representations of convex distribution-invariant risk measures. Examples of risk measures which are continuous on compacts are, in particular, average value at risk (AVaR) and shortfall risk. The industry standard value at risk (VaR) is not continuous on compacts, and the general large deviation results do not apply. Nevertheless, also in the case of value at risk we derive upper large deviation bounds – this time by direct calculations.

In Section 1.3 we return to risk measures which are continuous on compacts and investigate the rate function of large deviations of risk measurements in a special case. That is, actual risk measurements are based on empirical distributions of samples of the true distribution. In this situation, the rate function can be determined for different dependence structures of the simulated observations, i.e. independent or Markovian samples. In particular, if simulations are made independently, the rate function equals the minimal relative entropy under a risk measure constraint.

Based on general methods of Csiszar (1975), we calculate this minimal relative entropy explicitly for both average value at risk and shortfall risk, cf. Sections 1.4 and 1.5. In the first case, the analysis uses a particular representation of average value at risk as the expected loss under the worst case measure which can be computed by means of the Neyman-Pearson lemma. For average value at risk, the constraint set of the minimization problem is in general not convex, and the calculation is quite involved. Necessary and sufficient conditions for the existence of a solution are formulated in terms of the parameters of the problem. For AVaR, we solve the original problem in two steps. The first step consists of minimizing the relative entropy under a linear constraint. Minimizing densities and minimal relative entropies are explicitly calculated. In a second step, a minimization problem with three varying parameters has to be solved. In Section 1.5 we finally study the minimization of the relative entropy under a shortfall risk constraint. In contrast to average value at risk, this problem involves only a linear constraint.

The second chapter investigates dynamic risk measurements. Section 2.2 provides an axiomatic framework for a class of dynamic risk measures. We focus on risk measures of dynamic cash flows in discrete time with a finite time horizon. By assumption, acceptability of terminal positions depends on their conditional distribution only. A representation theorem in terms of distribution-invariant static risk measures is proved in Section 2.3. For this purpose it turns out to be useful to interpret distribution-invariant static risk measures as functionals on a space of real probability measures. Properties of such functionals are derived in Section 2.3.1.

The suggested axiomatic framework for dynamic risk measures does not imply any consistency of risk measurements at different points in time. In Section 2.4 we suggest therefore two notions of dynamic consistency, and prove that consistency and properties of representing static risk measures are closely related. The concept of measure convex sets known from Choquet theory is key to the complete characterization of static risk measures that correspond to consistent dynamic risk measures.

Measure convexity can be interpreted using the concept of compound lotteries. This leads to a characterization of a class of static risk measures closely connected to shortfall risk which is one of the main results of Chapter 2, see Theorem 2.5.3. Finally, this result is applied to dynamically consistent, convex risk.

## Part II – Models of Financial Risk

While the first part of this thesis studies how risk can be quantified, in the second part we investigate specific models of financial risk. In two case studies, we analyze two aspects of the microstructure of financial markets. In Chapter 3 we propose a model of firms which interact with their business partners and analyze the implications for credit risk in large loan portfolios. Chapter 4 deals in contrast with market interaction of investors with bounded rationality. An evolutionary market selection mechanism and the long-run performance of strategies are here the main focus.

Credit contagion refers to the propagation of economic distress from one firm to another. Using methods from the theory of interacting particle systems, Chapter 3 proposes a model for these contagion phenomena, assuming they are due to the local interaction of firms in a business partner network. Firms can be in two states, "low liquidity" and "high liquidity." The dynamics of the liquidity states of firms is described by a voter process. We assume that the structure of the business partner network can be described by the $d$-dimensional lattice $\mathbb{Z}^d$. For the voter model on $\{0,1\}^{\mathbb{Z}^d}$ a complete convergence theorem that characterizes the long-run behavior is well known, cf. Liggett (1985). For $d > 2$ we establish in Theorem 3.3.7 a refinement which is useful in order to separate systematic from contagion risk.

In a next step, we study aggregate credit losses on large portfolios of financial positions contracted with firms subject to credit contagion. In Section 3.4 we provide an explicit Gaussian approximation of the distribution of portfolio losses. Due to the interaction of firms, the re-scaling in this central limit theorem is non-standard. We find that the contagion process has typically a second order effect on portfolio losses. It induces additional fluctuations of losses around their averages, whose size depends on the denseness $d$ of the business partner network.

In Chapter 4 we investigate an evolutionary market selection model. The axiom of profit maximization or rationality of agents is key to neoclassical economics. Often it is justified by the market selection hypothesis, which argues that maximization is the long-run market behavior induced by an evolutionary selection process. This argument seems to be intuitively appealing, but needs a rigorous analysis.

An explicit model for a market selection mechanism has been proposed by Blume and Easley (1992). In an asset model with endogenous prices in discrete time, agents follow simple trading strategies. They keep the proportion of wealth invested in each asset fixed over time and reinvest their payoffs. The market process induces a redistribution of wealth among traders. Blume and Easley (1992) investigate the long-run behavior of the selection process. Under strong conditions on the underlying random variables and the payoff structure of the assets they identify the unique survivor of the market selection process.

In Chapter 4 we provide a continuous time approximation of the model of Blume and Easley (1992) for general payoffs of the assets. In Section 4.3.1 we prove a functional limit theorem for the wealth process which is closely related to the well known Euler scheme. In Section 4.3.2 conditions on the payoff structure of the assets are identified that guarantee convergence. We suggest an economically meaningful model for the dividend processes and their convergence. The notion of locally finite kernels turns out to be useful.

The continuous time approximation of the wealth process is given by the solution of an integral equation in a random environment. For constant asset returns, the integral equation reduces to an autonomous ordinary differential equation. In Section 4.4 we analyze its long-run asymptotic behavior using techniques related to Lyapunov functions. Finally, we compare our results to the benchmark of profit-maximizing investors. These equilibrium solutions are closely related to the asymptotic behavior of the evolutionary model.

## Acknowledgements

It is a great pleasure to gratefully acknowledge the support of my advisor Hans Föllmer. He promoted an interdisciplinary environment which I found extremely inspiring for my research. His instructive way to discuss both economic and mathematical topics provided a sound guideline and has been a continual source of fascination for the subject. His suggestions and inspiring remarks significantly enhanced the content and improved the presentation of this thesis.

A very special thanks also goes to Boris Buchmann and Kay Giesecke for fruitful collaborations leading to results which form the basis of the second part of this thesis, cf. Giesecke and Weber (2004a), Giesecke and Weber (2004b), and Buchmann and Weber (2004a). Kay drew my attention to credit risk models and the problem of credit contagion. He considerably influenced my view on credit risk and contributed to the clear presentation and economic interpretation of the results in Chapter 3. Chapter 4 grew out of discussions with Boris in Berlin, Hannover and München. I would also like to thank Claudia Klüppelberg for inviting me twice to TU München, and Ludwig Baringhaus and Rudolf Grübel for the opportunity to work on the project at the Institut für mathematische Stochastik in Hannover in December 2003. I wish to thank Boris Buchmann, Rüdiger Frey, Rüdiger Kiesel, Torsten Kleinow, Claudia Klüppelberg, Christoph Kühn, Thilo Liebig, András Löffler, Xuerong Mao, and Ronald Meester for inviting me to present my work at research seminars.

Moreover, I would like to thank my friends and colleagues Stefan Ankirchner, Peter Arts, Peter Bank, Christian Bender, Bernd Buchwald, Sonja Deppisch, Steffen Dereich, Felix Esche, Stefanie Fischer, Mihai Giurcanu, Anne Gundel, Matthias an der Heiden, Ulrich Horst, Pierre-Yves Louis, Markus Mildenstein, Matthias Müller, Irina Penner, Markus Riedle, Alexander Schied, Michel Sortais, Sina Tutsch, Christian Vormoor, Marc Christian Weber, and Wiebke Wittmüß. Bernd, Boris, Kay, Sina, and Steffen gave useful comments on parts of the final version of the thesis. Financial support of Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 373, "Quantification and Simulation of Economic Processes," and through Graduiertenkolleg 251, "Stochastic Processes and Probabilistic Analysis," is gratefully acknowledged.

Finally, I wish to express my gratitude to my parents for all their love and support.

# Part I

# Measures of Financial Risk

# Chapter 1

# Distribution-Invariant Risk Measures, Entropy, and Large Deviations

## 1.1 Introduction

The portfolios of banks consist of financial assets such as stocks, bonds, credits and options. The quantification of the risk associated with these positions is of crucial importance, since banks need to manage their risks and are obliged to respect regulatory constraints. This requires both suitable models of portfolio holdings and appropriate numerical measures of risk.

In practice, financial positions are frequently modelled as real-valued random variables on some underlying probability space. In such a setting, the modelling assumptions determine in particular the distributions of positions. A standard approach to measure financial risk is to use certain functionals of these distributions, namely static distribution-invariant risk measures.

A theory of such risk measures is already well developed. Nevertheless, the implementation of risk measurements requires additional analysis. Model distributions are often not directly tractable, but can only be simulated by Monte Carlo methods. If then risk measures are evaluated for the simulated instead of the model-implied distributions, actual risk measurements deviate from the model-implied risk and the errors of these measurements need to be analyzed.

In the current chapter we employ the theory of large deviations to study these errors for various risk measures. We investigate large deviation bounds for the industry standard value at risk (VaR) and a broad class of other static risk measures. VaR has certain drawbacks: value at risk does in general not encourage diversification of positions and neglects the size of large losses. For this reason, other risk measures with better properties have been proposed. Specific examples include robust mixtures of average value at risk, and shortfall risk. An axiomatic analysis of coherent risk measures was initiated in the seminal paper by Artzner et al. (1999) and later extended to general probability spaces and convex risk measures, see e.g. Delbaen (2002), Föllmer and Schied (2002b), Frittelli and Rosazza (2002), and Föllmer and Schied (2004).

The chapter is outlined as follows. In a first step, we describe how the error of the risk measurements and large deviations are related. In Section 1.2 we investigate conditions under which a large deviation principle (LDP) holds for risk measurements. A LDP can be derived from a contraction principle, if the underlying risk measure satisfies a certain regularity property, i.e., is continuous on compacts. This notion is introduced in Section 1.2.1, and a contraction principle for the corresponding class of risk measures is formulated. Section 1.2.2 analyzes therefore the notion of continuity on compacts. In particular, we prove sufficient conditions in terms of robust representations of convex distribution-invariant risk measures. Examples of risk measures which are continuous on compacts are, in particular, average value at risk (AVaR) and shortfall risk. In contrast, the industry standard VaR is not continuous on compacts, and the general theory cannot be employed to control the error probabilities. Nevertheless, we provide an upper large deviation bound for VaR in Section 1.2.3.

In Section 1.3 we return to the large deviations of risk measurements based on risk measures which are continuous on compacts, and study the special case of empirical distributions. This allows us to refine our results and to characterize the rate function of the LDP more explicitly. Suppose that a financial institution wants to evaluate the risk of a financial position $X$ and generates samples of $X$ by some simulation procedure. Then, the decay of the error probabilities of the risk measurements is determined by the dependence structure of the samples. We investigate two cases: independent and Markovian observations. For independent samples the rate function of the large deviations of the risk measurements can be characterized as the minimal relative entropy under a risk measure constraint. For Markovian observations the rate function is given by the minimal value of a Fenchel-Legendre transform under a risk measure constraint.

Based on general methods of Csiszar (1975), we calculate the minimal relative entropy explicitly for both average value at risk and shortfall risk in Sections 1.4 and 1.5. In the case of AVaR, the analysis uses a particular representation of AVaR as the expected loss under the worst case measure which can be computed by means of the

Neyman-Pearson lemma, see equation (1.27). The constraint set of the minimization problem is in general not convex, and the calculation is quite involved. Necessary and sufficient conditions for the existence of a solution are formulated in terms of the parameters of the problem. For AVaR, we solve the original problem in two steps. The first step consists of minimizing the relative entropy under a linear constraint. Minimizing densities and minimal relative entropies are explicitly calculated. In a second step, a minimization problem with three varying parameters has to be solved. In the final section, we consider the entropy minimization problem under a shortfall risk constraint. In contrast to average value at risk, the calculation of the minimal relative entropy only involves a linear constraint.

## 1.2 Large Deviation Bounds

In the current section we investigate large deviations for risk measurements of financial positions. In Section 1.2.1 we will introduce the class of risk measures which are continuous on compacts. For these risk measures a large deviation principle can be derived from the contraction principle, and the rate function is characterized by a variational principle. Section 1.2.2 provides examples of risk measures which are continuous on compacts. An upper large deviation bound for *value at risk* will be derived in Section 1.2.3. Since this industry standard is not continuous on compacts, the general theory of Section 1.2.1 does not apply in this case.

### 1.2.1 Continuity on Compacts and a Contraction Principle

In this section we investigate under which conditions a contraction principle applies to distribution-invariant risk measures. We always assume that $(\Omega, \mathcal{F}, P)$ is a rich probability space, i.e. a probability space on which a random variable with continuous distribution exists. We recall the following definition.

***Definition 1.2.1.*** *A mapping $\rho : L^\infty(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is called a distribution-invariant risk measure, if it satisfies the following conditions for all $X, Y \in L^\infty$:*

- Monotonicity: *If $X \leq Y$, then $\rho(X) \geq \rho(Y)$.*
- Translation-invariance: *If $m \in \mathbb{R}$, then $\rho(X + m) = \rho(X) - m$.*
- Distribution-invariance: *If $P \circ X^{-1} = P \circ Y^{-1}$, then $\rho(X) = \rho(Y)$.*

We denote by $\mathcal{M}_{1,c} = \mathcal{M}_{1,c}(\mathbb{R})$ the space of Borel probability measures on $\mathbb{R}$ with compact support. A distribution-invariant risk measure $\rho$ defines a functional $\rho'$ :

$\mathcal{M}_{1,c} \to \mathbb{R}$ by $\rho'(\mu) = \rho(X)$ for some $X \in L^\infty$ with distribution $\mathcal{L}(X) := P \circ X^{-1} = \mu$. For more details see Chapter 2.

We consider the following situation. Assume that we are interested in the risk of a financial position $X \in L^\infty$ with distribution $\mu = \mathcal{L}(X)$. Suppose that the distribution $\mu$ is not directly tractable, but that samples $\mu$ can be generated.

For example, let $(X_i)$ be a sequence of independent random variables on the probability space $(\Omega, \mathcal{F}, P)$ with identical distribution $\mu$. The empirical distribution of the first $n$ samples $X_1, X_2, \ldots, X_n$ is then given by the random measure

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}.$$

Here, $\delta_x$ denotes the Dirac measure placing all mass on $x \in \mathbb{R}$. Then $(\mu_n)$ converges $P$-almost surely to $\mu$ in the weak topology.

A naive Monte Carlo procedure for simulating $\rho(X)$ is to calculate $\rho'(\mu_n)$, $n \in \mathbb{N}$. We will provide a simple regularity condition which ensures $\rho'(\mu_n) \to \rho(X)$ P-almost surely as $n \to \infty$. A possible measure of the quality of the $n$th approximation is the probability that the error of the simulated risk deviates from the true risk of $X$ by more than a given bound $\epsilon > 0$, i.e.

$$P\left(\left|\rho'(\mu_n) - \rho(X)\right| > \epsilon\right).$$

Under suitable conditions we will derive asymptotic upper and lower exponential bounds for these error probabilities.

We recall the notion of a large deviations principle. $\mathcal{X}$ always denotes a topological space.

***Definition 1.2.2.*** *The function $I : \mathcal{X} \to [0, \infty]$ is called a* rate function *if*

*(1) $I \not\equiv \infty$.*

*(2) $I$ is lower semicontinuous.*

*(3) The level sets $\{I \le c\}$, $c \in \mathbb{R}$, are compact.*

If $\Gamma \subseteq \mathcal{X}$, we write $I(\Gamma) = \inf_{x\in\Gamma} I(x)$. $\bar{\Gamma}$ denotes the closure and $\Gamma^\circ$ the open interior of $\Gamma$.

***Definition 1.2.3.*** *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\mathcal{B}$ be a $\sigma$-algebra on $\mathcal{X}$. For $n \in \mathbb{N}$, let*

$$\chi_n : (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{B})$$

*be measurable. The sequence of random elements $\chi_n$ of $\mathcal{X}$ satisfies a* large deviation principle (LDP) *with rate $(\gamma_n)$ and rate function $I$, if the following conditions hold for all $\Gamma \in \mathcal{B}$:*

$$\limsup_{n\to\infty} \frac{1}{\gamma_n} \log P\left(\chi_n \in \Gamma\right) \quad \leq \quad -I(\bar{\Gamma}), \tag{1.1}$$

$$\liminf_{n\to\infty} \frac{1}{\gamma_n} \log P\left(\chi_n \in \Gamma\right) \quad \geq \quad -I(\Gamma^\circ). \tag{1.2}$$

On the space $\mathcal{M}_{1,c}$ we will consider different topologies, namely the weak topology and the $\tau$-topology. If nothing else is said, $\mathcal{M}_{1,c}$ will be endowed with the weak topology. For any risk measure which is regular enough, a LDP for risk measurements can easily be derived from a LDP for distributions, e.g. Sanov's Theorem. More specifically, if the risk measure is continuous on compacts, we only need to apply a contraction principle.

***Definition 1.2.4.*** *A distribution-invariant risk measure $\rho : L^\infty \to \mathbb{R}$ is called* continuous on compacts*, if for all compact sets $K \subseteq \mathbb{R}$ the restriction of $\rho'$ to $\mathcal{M}_1(K)$ is continuous. Here, $\mathcal{M}_1(K)$ denotes the space of probability measures supported in $K$.*

For risk measures which are continuous on compacts a LDP is a simple consequence of the contraction principle.

***Proposition 1.2.5.*** *Let $\rho$ be a distribution-invariant risk measure that is continuous on compacts. Assume that $(\mu_n) \subseteq \mathcal{M}_{1,c}$ is a sequence of random measures that satisfies a LDP with rate $(\gamma_n)$ and rate function $I$. Additionally, assume that there exists a compact set $K \subseteq \mathbb{R}$ such that supp $\mu_n \subseteq K$ for all $n$. Then $(\rho'(\mu_n))_n$ satisfies a LDP with rate $(\gamma_n)$ and rate function*

$$J(x) := \inf \left\{ I(\nu) : \nu \in \mathcal{M}_{1,c},\ x = \rho'(\nu) \right\}.$$

*Proof.* The proposition is a direct consequence of the contraction principle for Hausdorff spaces (see Dembo and Zeitouni (1998), Theorem 4.2.1, p. 126). □

The notion of continuity on compacts is weaker than global continuity of risk measures with respect to the weak topology. Before we characterize continuity on compacts in the next section, we demonstrate that continuity on the whole space cannot be expected.

***Example 1.2.6.*** *Let $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}_{[0,1]}, \mathrm{unif}[0,1])$. The functional $\rho : X \mapsto -\int X dP$ defines a distribution-invariant convex risk measure on $L^\infty(\Omega, \mathcal{F}, P)$ which is continuous on compacts. But $\rho$ is not continuous:*

*Define $X \equiv 0$, and*

$$X_n(\omega) = \begin{cases} -n^2 & \text{if} \quad \omega \in [0, 1/n], \\ 0 & \text{else.} \end{cases} \tag{1.3}$$

*Let $\mu_n := \mathcal{L}(X_n)$ and $\mu := \mathcal{L}(X)$. Obviously, $\mu_n$ converges weakly to $\mu$ as $n \to \infty$, but*

$$\rho'(\mu_n) = \rho(X_n) = n \stackrel{n\to\infty}{\longrightarrow} \infty \neq 0 = \rho(X) = \rho'(\mu). \tag{1.4}$$

### 1.2.2 Examples

In this section we will provide examples of risk measures which are continuous on compacts and thus satisfy the contraction principle of the preceding section. The industry standard *value at risk* is not continuous on compacts. Since the contraction principle is thus not applicable, we will investigate lower and upper large deviation bounds for value at risk in the separate Section 1.2.3.

The following proposition is elementary, but allows us to identify examples of risk measures which are continuous on compacts. We recall that a risk measure $\rho$ is called *continuous from above*, if $X_n \searrow X$ $P$-almost surely implies $\rho(X_n) \nearrow \rho(X)$. Analogously, $\rho$ is called *continuous from below*, if $X_n \nearrow X$ $P$-almost surely implies $\rho(X_n) \searrow \rho(X)$.

***Proposition 1.2.7.*** *Let $\rho$ be a distribution-invariant risk measure. The following conditions are equivalent:*

*(1) $\rho$ is continuous on compacts.*

*(2) $\rho$ is both continuous from above and from below.*

*(3) $\rho$ is continuous for bounded sequences, i.e. for every bounded sequence $(X_n)$ converging $P$-almost surely to some $X$ it holds that $\lim_{n\to\infty} \rho(X_n) = \rho(X)$.*

*Proof.* (2)⇒(3): Let $(X_n)$ be bounded and converging $P$-almost surely to $X$. Then $\left(\sup_{m\geq n} X_m\right)_n$ and $(\inf_{m\geq n} X_m)_n$ converge to $X$ from above and below, respectively. Thus,

$$\rho(X) = \lim_n \rho\left(\sup_{m\geq n} X_m\right) \leq \liminf_{n\to\infty} \rho(X_n) \leq \limsup_{n\to\infty} \rho(X_n) \leq \lim_n \rho\left(\inf_{m\geq n} X_n\right) = \rho(X).$$

(3)⇒(2): If $(X_n)$ converges to $X$ from below or from above, then $(X_n)$ is bounded. This implies the claim.

(1)⇒(3): Let $(X_n)$ be a bounded sequence converging $P$-almost surely to some $X$. Then there exists a compact $K \subseteq \mathbb{R}$ such that $P$-almost surely $X_n, X \in K$ $(n \in \mathbb{N})$. Clearly,

$\mathcal{L}(X_n), \mathcal{L}(X) \in \mathcal{M}_1(K)$ and $\mathcal{L}(X_n) \to \mathcal{L}(X)$. Thus, $\rho(X_n) = \rho'(\mathcal{L}(X_n)) \to \rho'(\mathcal{L}(X)) = \rho(X)$.
(3)⇒(1): Let $K \subseteq \mathbb{R}$ be a compact set and assume that $\mu_n \Rightarrow \mu$ for $\mu_n, \mu \in \mathcal{M}_1(K)$. Denote by $F_n$, $F$ the distribution functions of $\mu_n$, $\mu$, respectively. Since $(\Omega, \mathcal{F}, P)$ is rich, there exists a random variable $Z$ with $\mathcal{L}(Z) = \text{unif}\,[0,1]$. Define $X_n := F_n^{-1}(Z)$, $X := F^{-1}(Z)$, where $F_n^{-1}$ and $F^{-1}$ are the right-continuous inverses of $F_n$ and $F$, respectively. Observe that $X_n \to X$ $P$-a.s. as $n \to \infty$. Moreover, $X_n, X \in K$ $P$-a.s. Hence,

$$\rho'(\mu_n) = \rho(X_n) \to \rho(X) = \rho'(\mu). \tag{1.5}$$

□

We provide examples for risk measures which are continuous on compacts. We will investigate both coherent and convex risk measures. The current industry standard *value at risk* is not continuous on compacts, as the following example shows.

***Example 1.2.8.*** *Value at risk at level $\lambda \in (0,1)$ is defined as*

$$VaR_\lambda(X) = \inf\left\{m \in \mathbb{R} : \ P[m + X < 0] \le \lambda\right\}.$$

*In order to see that value at risk is not continuous on compacts, we consider the following example. Let the probability space $(\Omega, \mathcal{F}, P)$ given by the unit interval $[0,1]$ with Lebesgue measure. For $\frac{1}{n} < 1 - \lambda$ define $X_n = \mathbf{1}_{[\lambda+1/n,1]}$. Then $X_n \nearrow X := \mathbf{1}_{[\lambda,1]}$ as $n \to \infty$. But, $VaR_\lambda(X_n) = 0$ does not converge to $VaR_\lambda(X) = -1$ as $n \to \infty$.*

For coherent risk measures we will state sufficient conditions for continuity on compacts. We recall that a risk measure is coherent, if it satisfies for $X, Y \in L^\infty$ both

- convexity: $\rho(\alpha X + (1-\alpha)Y) \le \alpha\rho(X) + (1-\alpha)\rho(Y)$, $\alpha \in [0,1]$, and
- positive homogeneity: $\rho(\alpha X) = \alpha\rho(X)$, $\alpha \ge 0$.

An important example of a distribution-invariant coherent risk measure is *average value at risk*.

***Example 1.2.9.*** Average value at risk *at level $\lambda \in (0,1]$ is defined as*

$$AVaR_\lambda(X) = \frac{1}{\lambda}\int_0^\lambda VaR_\gamma(X)d\gamma.$$

*Average value at risk is continuous on compacts.*

*Proof.* According to Theorem 4.39 of Föllmer and Schied (2002c), $AVaR_\lambda$ is continuous from below. By Theorem 4.26 in Föllmer and Schied (2002c), $AVaR_\lambda$ is continuous from above. Thus, $AVaR$ s continuous on compacts. □

Average value at risk is an important building block for coherent distribution-invariant risk measures. We quote the following theorem of Kusuoka (2001).

***Theorem 1.2.10.*** *On a rich probability space, a coherent distribution-invariant risk measure $\rho$ is continuous from above, if and only if*

$$\rho(X) = \sup_{\mu \in \mathcal{M}} \int_{(0,1]} AVaR_\lambda(X)\mu(d\lambda) \tag{1.6}$$

*for some set $\mathcal{M} \subseteq \mathcal{M}_1((0,1])$. Here, $\mathcal{M}_1((0,1])$ denotes the space of Borel probability measures on $(0,1]$.*

We denote by $AVaR_0(X) := \|X^-\|$ the essential infimum of $X$. Motivated by Kusuoka's Theorem, we introduce the following notation. If a measure $\mu \in \mathcal{M}_1([0,1])$, i.e. $\mu$ is a Borel probability measure on $[0,1]$, then we write

$$\rho_\mu(X) := \int_{[0,1]} AVaR_\lambda(X)\mu(d\lambda) \qquad (X \in L^\infty).$$

The following lemma will be useful.

***Lemma 1.2.11.*** *Let $\mu, (\mu_n) \subseteq \mathcal{M}_1((0,1])$. Assume that $(\mu_n)$ converges weakly to $\mu$. Let $X_n \nearrow X$. Then $\rho_{\mu_n}(X_n) \to \rho_\mu(X)$.*

*Proof.* For $\epsilon > 0$ define the continuous functions

$$\phi_\epsilon(\lambda) = \begin{cases} 1, & \lambda \le \epsilon, \\ 1 - \frac{\lambda - \epsilon}{\epsilon}, & \epsilon < \lambda < 2\epsilon, \\ 0, & \lambda \ge 2\epsilon. \end{cases}$$

We can estimate

$$\begin{aligned} u_n &:= \left| \int_{(0,1]} AVaR_\lambda(X_n) - AVaR_\lambda(X)\mu_n(d\lambda) \right| \\ &\le \sup_{\lambda \in [\epsilon,1]} |AVaR_\lambda(X_n) - AVaR_\lambda(X)| \; + \; 2(\|X_1\| + \|X\|) \cdot \int \phi_\epsilon(\lambda)\mu_n(d\lambda). \end{aligned}$$

The function $\lambda \mapsto |AVaR_\lambda(X_n) - AVaR_\lambda(X)| =: v_n(\lambda)$ converges pointwise to zero by continuity from below of $AVaR_\lambda$ ($\lambda \in (0,1]$). Since for any $\lambda \in (0,1]$ we have $AVaR_\lambda(X_n) \searrow AVaR_\lambda(X)$, we obtain $v_n(\lambda) \ge v_{n+1}(\lambda)$ ($n \in \mathbb{N}$). Let $\hat{\epsilon} > 0$. Thus,

$$M_n \; := \; \{\lambda \in [\epsilon,1] : \; v_m(\lambda) < \hat{\epsilon} \quad \forall m \ge n\} \; = \; \{\lambda \in [\epsilon,1] : \; v_n(\lambda) < \hat{\epsilon}\},$$

and $M_n \subseteq M_{n+1}$ ($n \in \mathbb{N}$). Since $v_n$ is continuous, $M_n$ is open. Pointwise convergence of $(v_n)$ to 0 implies $\bigcup M_n = [\epsilon, 1]$. By compactness of $[\epsilon,1]$, there exists $\hat{n} \in \mathbb{N}$ such that $\mathcal{M}_{\hat{n}} = [\epsilon, 1]$. Hence, $(v_n)$ converges uniformly to zero.

The sequence of integrals $\left(\int \phi_\epsilon(\lambda)\mu_n(d\lambda)\right)_n$ converges to $\int \phi_\epsilon(\lambda)\mu(d\lambda)$. Since $\mu\{0\} = 0$, we obtain

$$\lim_{\epsilon\to 0} \int \phi_\epsilon(\lambda)\mu(d\lambda) = 0.$$

This implies $\lim_{n\to\infty} u_n = 0$. Finally,

$$\begin{aligned} & \lim_{n\to\infty} \rho_{\mu_n}(X_n) \\ = \quad & \lim_{n\to\infty} \int_{(0,1]} \left(AVaR_\lambda(X_n) - AVaR_\lambda(X)\right)\ \mu_n(d\lambda) \quad + \quad \lim_{n\to\infty} \rho_{\mu_n}(X) \\ = \quad & \rho_\mu(X). \end{aligned}$$

□

***Proposition 1.2.12.*** *Suppose a coherent distribution-invariant risk measure $\rho$ admits a representation (1.6) for some weakly compact set $\mathcal{M} \subseteq \mathcal{M}_1((0,1])$. Then the supremum in (1.6) is actually a maximum, and $\rho$ is continuous on compacts.*

*Proof.* There exists a sequence $(\mu_n) \subseteq \mathcal{M}$ such that $\int_{(0,1]} AVaR_\lambda(X)\mu_n(d\lambda)$ converges to $\rho(X)$. Since $\mathcal{M}$ is weakly compact, we may assume that $(\mu_n)$ converges weakly to a limit in $\mathcal{M}$, say $\mu$. Thus,

$$\rho(X) = \lim_{n\to\infty} \int_{(0,1]} AVaR_\lambda(X)\mu_n(d\lambda) = \int_{(0,1]} AVaR_\lambda(X)\mu(d\lambda).$$

The last equation follows, since $\lambda \mapsto AVaR_\lambda(X)$ is continuous and bounded in $[-\|X\|, \|X\|]$. Hence, the supremum in (1.6) is actually a maximum.

By Theorem 1.2.10 $\rho$ is continuous from above. Let now $X_n \nearrow X$. It holds that

$$\rho(X_1) \geq \rho(X_2) \geq \cdots \geq \rho(X_n) \geq \cdots \geq \rho(X).$$

Clearly, $(\rho(X_n))$ converges to some $c \geq \rho(X)$. If for some subsequence $(X_{n_k})$ of $(X_n)$ the risks $\rho(X_{n_k})$ converge to $\rho(X)$ as $k \to \infty$, then $\rho$ is continuous from below. We show now that such a subsequence exists.

Since the supremum in (1.6) is attained, there exist $\mu, (\mu_n) \subseteq \mathcal{M}_1((0,1])$ such that $\rho(X) = \rho_\mu(X)$, $\rho(X_n) = \rho_{\mu_n}(X_n)$. Since $\mathcal{M}$ is weakly compact, we may choose a subsequence $(\mu_{n_k})$ of $(\mu_n)$ which converges to a limit in $\mathcal{M}$, say $\hat{\mu}$. Then by Lemma 1.2.11,

$$c = \lim_k \rho(X_{n_k}) = \rho_{\hat{\mu}}(X).$$

Finally, observe that by (1.6),

$$c = \int_{(0,1]} AVaR_\lambda(X)\hat{\mu}(d\lambda) \leq \rho(X).$$

Thus, $c = \rho(X)$. □

The condition of Proposition 1.2.12 is, of course, satisfied, if the set $\mathcal{M}$ is a singleton. In this case, $\rho$ is simply a mixture of average value at risk at different levels. By a theorem of Schmeidler (1986) the class of such risk measures is closely related to the family of distribution-invariant risk measures that are comonotonic additive.

***Definition 1.2.13.*** *Two measurable functions $X$ and $Y$ on $(\Omega, \mathcal{F}, P)$ are called $P$-*comonotone *if*

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geq 0$$

*for $P \otimes P$-almost all $(\omega, \omega') \in \Omega \times \Omega$.*

*A risk measure $\rho : L^\infty \to \mathbb{R}$ is* comonotonic additive, *if*

$$\rho(X + Y) = \rho(X) + \rho(Y)$$

*whenever $X$ and $Y$ are $P$-comonotone.*

***Theorem 1.2.14.*** *On a rich probability space, the class of risk measures*

$$\rho_\mu(X) = \int AVaR_\lambda(X)\mu(d\lambda), \qquad \mu \in \mathcal{M}_1([0,1])$$

*is precisely the class of all distribution-invariant convex risk measures on $L^\infty$ that are comonotonic additive. In particular, these risk measures are also coherent.*

*If $\mu\{0\} = 0$, then $\rho_\mu$ is continuous on compacts.*

*Proof.* For the first part of the theorem see Theorem 4.65 in Föllmer and Schied (2004). Moreover, since $AVaR_\lambda$ $(\lambda \in (0,1])$ is continuous from above and below, the same holds for any mixture $\rho_\mu$, if $\mu\{0\} = 0$. The claim follows from Proposition 1.2.7. □

For general convex risk measures a necessary condition for continuity on compacts can be given in terms of level sets of a penalty function. The criterion is based on the following robust representation theorem.

***Theorem 1.2.15.*** *On a rich probability space, distribution-invariant convex risk measures are continuous from above, if and only if*

$$\rho(X) = \sup_{\mu \in \mathcal{M}_1((0,1])} \left( \int_{(0,1]} AVaR_\lambda(X)\mu(d\lambda) - \beta(\mu) \right). \tag{1.7}$$

*Here, $\beta : \mathcal{M}_1((0,1]) \to \mathbb{R} \cup \{\infty\}$ is a penalty function with $\inf_{\mu \in \mathcal{M}_1((0,1])} \beta(\mu) \in \mathbb{R}$.*

***Proposition 1.2.16.*** *Let $\rho$ be represented as in (1.7). For $c \in \mathbb{R}$ define the level sets $\Lambda_c = \{\mu \in \mathcal{M}_1((0,1]) : \ \beta(\mu) \leq c\}$. Assume that $\Lambda_c$ is weakly compact for all $c \in \mathbb{R}$. Then the supremum in (1.7) is actually a maximum. Let*

$$\mathcal{S} := \{\mu \in \mathcal{M}_1((0,1]) : \ \exists X \in L^\infty \text{ with } \rho(X) = \rho_\mu(X) - \beta(\mu)\}.$$

*If some sequence* $(\mu_n) \subseteq \mathcal{M}_1((0,1])$ *converges to* $\mu \in \mathcal{S}$ *as* $n \to \infty$, *then* $\beta(\mu) = \liminf_{n\to\infty} \beta(\mu_n)$. *Moreover,* $\rho$ *is continuous on compacts.*

*Proof.* By (1.7) there exists a sequence $(\mu_n) \subseteq \mathcal{M}$ such that $(\rho_{\mu_n}(X) - \beta(\mu_n))$ converges to $\rho(X)$. Observe that $|\rho_{\mu_n}(X)| \leq \|X\|$. Thus, convergence of $(\rho_{\mu_n}(X) - \beta(\mu_n))$ to $\rho(X)$ implies that $(\beta(\mu_n))$ is bounded from above by some $c \in \mathbb{R}$. Hence, $(\mu_n) \subseteq \Lambda_c$. By the weak compactness of $\Lambda_c$, we may assume that $(\mu_n)$ is convergent with limit $\mu \in \mathcal{M}$. Since $\lambda \mapsto AVaR_\lambda(X)$ is continuous and bounded, we obtain that $\lim_{n\to\infty} \rho_{\mu_n}(X) = \rho_\mu(X)$. Thus, also $(\beta(\mu_n))$ converges as $n \to \infty$ and

$$\rho(X) = \rho_\mu(X) - \lim_{n\to\infty} \beta(\mu_n).$$

Since $\beta$ is lower semicontinuous by assumption, it holds that $\lim_{n\to\infty} \beta(\mu_n) \geq \beta(\mu)$. Suppose that $\lim_{n\to\infty} \beta(\mu_n) > \beta(\mu)$. Then

$$\rho(X) = \rho_\mu(X) - \lim_{n\to\infty} \beta(\mu_n) < \rho_\mu(X) - \beta(\mu) \leq \rho(X),$$

a contradiction. Thus, $\rho(X) = \rho_\mu(X) - \beta(\mu)$.

Now let $(\mu_n)$ be any sequence converging to $\mu$. Then by lower semicontinuity of $\beta$, we get $\liminf_{n\to\infty} \beta(\mu_n) \geq \beta(\mu)$. Suppose that $\liminf_{n\to\infty} \beta(\mu_n) > \beta(\mu)$. Then

$$\rho(X) \leq \rho_\mu(X) - \liminf_{n\to\infty} \beta(\mu_n) < \rho_\mu(X) - \beta(\mu) \leq \rho(X),$$

a contradiction. Thus, $\beta(\mu) = \liminf_{n\to\infty} \beta(\mu_n)$.

By Theorem 1.2.15 $\rho$ is continuous from above. Let now $X_n \nearrow X$,. Then there exist $\mu, (\mu_n) \subseteq \mathcal{M}_1((0,1])$ with

$$\begin{aligned} \rho(X_n) &= \rho_{\mu_n}(X_n) - \beta(\mu_n), \\ \rho(X) &= \rho_\mu(X) - \beta(\mu). \end{aligned}$$

It holds that

$$\rho(X_1) \geq \rho(X_2) \geq \cdots \geq \rho(X_n) \geq \cdots \geq \rho(X).$$

Clearly, $(\rho(X_n))$ converges to some $c \geq \rho(X)$. If for some subsequence $(X_{n_k})$ of $(X_n)$ the risks $(\rho(X_{n_k}))$ converge to $\rho(X)$ as $k \to \infty$, then $\rho$ is continuous from below. We show now that such a subsequence exists.

Observe first that $|\rho_{\mu_n}(X_n)| \leq \|X_n\|$. Thus,

$$|\beta(\mu_n)| \leq |\rho(X_n)| + |\rho_{\mu_n}(X_n)| \leq |\rho(X_1)| + |\rho(X)| + \|X_1\| + \|X\|.$$

This implies that there exists $c \in \mathbb{R}$ such that $\mu_n \in \Lambda_c$ $(n \in \mathbb{N})$. Since $\Lambda_c$ is weakly compact, we may choose a subsequence $(\mu_{n_k})$ of $(\mu_n)$ which converges to a limit in $\mathcal{M}$, say $\hat{\mu}$. Then by Lemma 1.2.11,

$$\begin{aligned} c &= \lim_k \rho(X_{n_k}) = \lim_k \left( \rho_{\mu_{n_k}}(X_{n_k}) - \beta(\mu_{n_k}) \right) = \rho_{\hat{\mu}}(X) - \lim_k \beta(\mu_{n_k}) \\ &= \lim_k \left( \rho_{\mu_{n_k}}(X) - \beta(\mu_{n_k}) \right) \leq \rho(X). \end{aligned}$$

Thus, $c = \rho(X)$. □

Finally, we discuss an important class of distribution-invariant convex risk measures which are continuous on compacts, namely shortfall risk. Shortfall risk admits a representation of the form (1.7). But instead of employing this representation as a basis for the analysis, it will be more convenient to work with the following definition.

***Definition 1.2.17.*** *Let $\ell : \mathbb{R} \to \mathbb{R}$ be a convex loss function, i.e. an increasing, non constant and convex function. Assume that $z$ is an interior point of the range of $\ell$. We define the acceptance set*

$$\mathcal{A} = \left\{ X \in L^\infty : \int \ell(-X) dP \leq z \right\}.$$

*The shortfall risk is defined by*

$$\rho(X) = \inf \{ m \in \mathbb{R} : \ X + m \in \mathcal{A} \}.$$

Shortfall risk is a distribution-invariant risk measure which is continuous from above and below, cf. Proposition 4.59 & Theorem 4.26 in Föllmer and Schied (2002c). Thus, by Proposition 1.2.7 shortfall risk is continuous on compacts. Like average value at risk, shortfall risk has many desirable properties. In contrast to value at risk, it encourages diversification, since it is convex, and does not neglect the size of losses.

As discussed in Section 1.2.1, shortfall risk induces a functional $\rho' : \mathcal{M}_{1,c} \to \mathbb{R}$. If we denote by $\mathcal{N} = \{\mu \in \mathcal{M}_{1,c} : \ \rho'(\mu) \leq 0\}$ the acceptance set of shortfall risk on the level of distributions, and by $\mathcal{N}^c$ the corresponding rejection set, we can state the following interesting result. If $\mu, \nu \in \mathcal{N}$ are acceptable lotteries, then for any $\alpha \in [0,1]$ also the compound lottery $\alpha\mu + (1-\alpha)\nu$ is acceptable, i.e. an element of $\mathcal{N}$. Moreover, if $\nu, \mu \in \mathcal{N}^c$ are rejected lotteries, then for any $\alpha \in [0,1]$ also the compound lottery $\alpha\mu + (1-\alpha)\nu \in \mathcal{N}^c$. It can be shown that shortfall risk is essentially the only convex and distribution-invariant risk measure with these properties. For this reason, shortfall risk is closely related to dynamic risk measurements, cf. Chapter 2.

### 1.2.3 Large Deviation Bounds for $VaR$

In contrast to the examples of the preceding section the industry standard value at risk is not continuous on compacts. Proposition 1.2.5 is thus not applicable in this case. Although we do not obtain a full LDP, we will derive an upper large deviation bound for $VaR_\lambda$ ($\lambda \in (0,1)$) in the current section.

For this purpose it turns out to be useful to employ a different topology on the space of probability distributions than the weak topology, namely the $\tau$-topology. This topology is introduced in the next definition. If we use the $\tau$-topology instead of the weak topology, we will always state this explicitly.

***Definition 1.2.18.*** *Let $\mathcal{M}_1(\mathbb{R})$ be the set of probability measures on $\mathbb{R}$. We denote the family of bounded measurable functions by $B(\mathbb{R})$. The initial topology generated by the functionals $\mathcal{M}_1 \to \mathbb{R}$, $\mu \mapsto \int f d\mu$ $(f \in B(\mathbb{R}))$ is called the* $\tau$-topology.

***Remark 1.2.19.*** *The $\tau$-topology and the weak topology are related to each other as follows:*

*(1) Since $C_b(\mathbb{R}) \subseteq B(\mathbb{R})$, the $\tau$-topology is finer than the weak topology. Here, $C_b(\mathbb{R})$ denotes the set of continuous, bounded functions on $\mathbb{R}$. In particular, convergence in the $\tau$-topology implies convergence in the weak topology.*

*(2) The fact that $\mathbb{R}$ is a Polish space implies that the $\sigma$-algebra on $\mathcal{M}_1(\mathbb{R})$ generated by the $\tau$-topology equals the $\sigma$-algebra generated by the weak topology.*

We define the term *$\tau$-continuous on compacts* analogously to Definition 1.2.4. Even with respect to this notion $VaR$ is not continuous as the following example demonstrates. Thus, $VaR$ does not allow for a direct application of a contraction principle, even if a LDP for probability measures with respect to the $\tau$-topology is available.

As every distribution-invariant risk measure, $VaR_\lambda$ ($\lambda \in (0,1)$) induces a functional $\rho' : \mathcal{M}_{1,c} \to \mathbb{R}$. For simplicity, we write again $VaR_\lambda$ instead of $\rho'$.

***Example 1.2.20.*** *Let $\lambda \in (0,1)$. We define a sequence $(\mu_n)$ of measures by*

$$\mu_n = \left(\lambda + \frac{1}{n}\right)\delta_{\{0\}} + \left(1 - \lambda - \frac{1}{n}\right)\delta_{\{1\}}.$$

*Here, we assume that $1/n < 1 - \lambda$. Clearly, $(\mu_n)$ converges to $\mu = \lambda\delta_{\{0\}} + (1-\lambda)\delta_{\{1\}}$. in the $\tau$-topology. But $VaR_\lambda(\mu_n) = 0$ does not converge to $VaR_\lambda(\mu) = -1$ as $n \to \infty$.*

$VaR_\lambda$ ($\lambda \in (0,1)$) is not continuous on compacts, neither with respect to the weak nor to the $\tau$-topology. In general, we thus cannot approximate $VaR_\lambda(X)$ for

$X \in L^\infty$ by a naive Monte Carlo procedure. To be more precise: Let $K \subseteq \mathbb{R}$ be compact. If $(\mu_n) \subseteq \mathcal{M}_1(K)$ converges to $\mu \in \mathcal{M}_1(K)$ in the weak or the $\tau$-topology, then $(VaR_\lambda(\mu_n))$ does i.g. not converge to $VaR_\lambda(X)$. Nevertheless, we will obtain convergence, if $\mu$ possesses a unique $\lambda$-quantile.

***Definition 1.2.21.*** *Let $\mu \in \mathcal{M}_1(\mathbb{R})$ and $\lambda \in (0,1)$ be a given level. We will say that $\mu$ has a unique quantile at level $\lambda$, if the interval of $\lambda$-quantiles $[q_\lambda^-(\mu), q_\lambda^+(\mu)]$ is a singleton. Here, the bounds of the interval are given by*

$$q_\lambda^-(\mu) \;:=\; \inf\{y:\; \mu(-\infty,y] \geq \lambda\} \;=\; \sup\{y:\; \mu(-\infty,y) < \lambda\},$$

*and*

$$q_\lambda^+(\mu) \;:=\; \inf\{y : \mu(-\infty,y] > \lambda\} \;=\; \sup\{y : \mu(-\infty,y) \leq \lambda\}.$$

***Proposition 1.2.22.*** *If the sequence $(\mu_n) \subseteq \mathcal{M}_{1,c}$ converges to $\mu \in \mathcal{M}_{1,c}$ in the weak topology and $\mu$ has a unique quantile at level $\lambda$, then*

$$\lim_{n\to\infty} VaR_\lambda(\mu_n) = VaR_\lambda(\mu).$$

*Proof.* Since $\mu$ has a unique quantile $q$ at level $\lambda$, we get

(1) $\mu((-\infty,q']) < \lambda$ for all $q' < q$,

(2) $\mu((-\infty,q')) > \lambda$ for all $q' > q$.

We get therefore that

(1) $\limsup_{n\to\infty} \mu_n((-\infty,q']) < \lambda$ for all $q' < q$,

(2) $\liminf_{n\to\infty} \mu_n((-\infty,q')) > \lambda$ for all $q' > q$.

From (1) follows that for given $q' < q$ there exist $\epsilon > 0$ and a natural number $n_0$ such that for all integers $n$ larger than $n_0$ we have

$$\mu_n((-\infty,q']) \leq \lambda - \epsilon,$$

and thus $q_\lambda^+(\mu_n) > q'$. Since $q' < q$ is arbitrary, we obtain

$$\liminf_{n\to\infty} q_\lambda^+(\mu_n) \geq q.$$

Analogously, (2) implies

$$\limsup_{n\to\infty} q_\lambda^+(\mu_n) \leq q.$$

Hence, we have

$$\lim_{n\to\infty} VaR_\lambda(\mu_n) = -\lim_{n\to\infty} q_\lambda^+(\mu_n) = -q = VaR_\lambda(\mu).$$

□

As a consequence of the last proposition, a naive Monte Carlo approximation of $VaR_\lambda(X)$ for $X \in L^\infty$ and $\lambda \in (0,1)$ – as described in Section 1.2.1 – works $P$-almost surely, if $X$ has a unique quantile at level $\lambda$.

Finally, we provide an upper large deviation bound for value at risk.

***Proposition 1.2.23.*** *Assume that $(\mu_n) \subseteq \mathcal{M}_{1,c}$ is a sequence of random measures that satisfies a LDP with respect to the $\tau$-topology with rate $(\gamma_n)$ and rate function $I$. Then for given level $\lambda \in (0,1)$*

$$\limsup_{n\to\infty} \frac{1}{\gamma_n} \log P\left(|VaR_\lambda(\mu_n) - VaR_\lambda(\mu)| > \epsilon\right) \le -\left\{\inf_{x\ge\lambda} J_-^{\epsilon,\lambda}(x) \wedge \inf_{x\le\lambda} J_+^{\epsilon,\lambda}(x)\right\}$$

*where the rate function is given by*

$$J_\pm^{\epsilon,\lambda}(x) := \inf\{I(\nu) : \mu \in \mathcal{M}_{1,c}, \quad \nu((-\infty, -VaR_\lambda(\mu) \pm \epsilon]) = x\}.$$

*Proof.* First observe that for $x \in \mathbb{R}$

$$q_\lambda^+(\mu_n) < x \implies \mu_n((-\infty, x]) \ge \lambda, \tag{1.8}$$

$$q_\lambda^+(\mu_n) > x \implies \mu_n((-\infty, x]) \le \lambda. \tag{1.9}$$

This implies now that

$$\{VaR_\lambda(\mu_n) - VaR_\lambda(\mu) > \epsilon\} = \left\{q_\lambda^+(\mu_n) - q_\lambda^+(\mu) < -\epsilon\right\}$$

$$= \left\{q_\lambda^+(\mu_n) < -VaR_\lambda(\mu) - \epsilon\right\} \overset{(1.8)}{\subseteq} \{\mu_n((-\infty, -VaR_\lambda(\mu) - \epsilon]) \ge \lambda\}$$

Since the random measures $(\mu_n)$ satisfy a LDP with respect to the $\tau$-topology, the random variables $(\mu_n(-\infty, VaR_\lambda(\mu) - \epsilon]))_n$ satisfy a LDP with the same rate and rate function $J_-^{\epsilon,\lambda}$. Thus,

$$\limsup_{n\to\infty} \frac{1}{\gamma_n} \log P\left(VaR_\lambda(\mu_n) - VaR_\lambda(\mu) > \epsilon\right) \le -\inf_{x\ge\lambda} J_-^{\epsilon,\lambda}(x) \tag{1.10}$$

Similarly, we get from (1.9)

$$\{VaR_\lambda(\mu_n) - VaR_\lambda(\mu) < -\epsilon\} = \left\{q_\lambda^+(\mu_n) - q_\lambda^+(\mu) > \epsilon\right\}$$

$$= \left\{q_\lambda^+(\mu_n) > -VaR_\lambda(\mu) + \epsilon\right\} \overset{(1.9)}{\subseteq} \{\mu_n((-\infty, -VaR_\lambda(\mu) + \epsilon]) \le \lambda\}$$

Arguing like above, we obtain

$$\limsup_{n\to\infty} \frac{1}{\gamma_n} \log P\left(VaR_\lambda(\mu_n) - VaR'_\lambda(\mu) < -\epsilon\right) \le -\inf_{x\le\lambda} J_+^{\epsilon,\lambda}(x) \tag{1.11}$$

From (1.10) and (1.11) follows our claim. □

***Remark 1.2.24.*** *For specific sampling procedures, Fu et al. (2003) derive related upper bounds for quantile estimators in terms of Legendre transforms of cumulant generating functions.*

## 1.3 Rate Functions and Dependence of Observations

As we have seen in the last section, for risk measures which are continuous on compacts a LDP follows from a LDP of the underlying random measures. In the current section we specialize to the case of empirical measures. The rate function of the LDP can be determined for different dependence structures of the observations. We summarize well-known results for independent and Markovian observations, and apply them to the context of risk measurements.

In the case of independent observations, the rate function for the LDP of risk measurements can be calculated as the minimal relative entropy under a risk measure constraint. For average value at risk and shortfall risk we will explicitly solve this minimization problem in Sections 1.4 & 1.5.

In the whole section we consider the following situation. Let $(\Omega, \mathcal{F}, P)$ be a rich probability space, and let $X \in L^\infty$ be a financial position with distribution $\mu \in \mathcal{M}_{1,c}$. By $\rho : L^\infty \to \mathbb{R}$ we denote a distribution-invariant risk measure. We assume that a financial institution is interested in assessing the risk $\rho(X)$, but does not explicitly know the distribution $\mu$. Instead the institution can simulate identically distributed random variables $(X_i)_{i\in\mathbb{N}}$ with law $\mu$. For $n \in \mathbb{N}$ the institution observes the empirical measure

$$\mu_n \;=\; \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i},$$

and uses $\rho'(\mu_n)$ as an approximation of the true risk $\rho(X)$.

***Corollary 1.3.1.*** *If $\rho$ is a distribution-invariant risk measure that is continuous on compacts, then the following statements hold:*

*(1) If the sequence of empirical measures converges $P$-almost surely to $\mu$, then $P$-almost surely*

$$\lim_{n\to\infty} \rho'(\mu_n) = \rho'(\mu).$$

*(2) If the sequence of empirical measures satisfies a LDP with rate $(\gamma_n)$ and rate function $I$, then $(\rho'(\mu_n))$ satisfies a LDP with rate $(\gamma_n)$ and rate function*

$$J(x) = \inf\{I(\nu) : \; \nu \in \mathcal{M}_{1,c}, \; x = \rho'(\nu)\}.$$

*If $J$ has a unique zero, then $P$-a.s. $\lim_{n\to\infty} \rho'(\mu_n) = \rho'(\mu)$.*

*Proof.* The first statement follows simply from continuity on compacts, since all empirical measures $\mu_n$ are concentrated on a compact set. The second statement is a direct application of Proposition 1.2.5. □

**Independent variables** If the observations of the financial institution are made independently, the empirical measures $(\mu_n)$ converge $P$-almost surely to $\mu$. Thus, a strong law of large numbers holds for the risk measurements. At the same time, we have the following LDP for the risk measures:

***Proposition 1.3.2.*** *Let $\rho$ be continuous on compacts. Then $(\rho'(\mu_n))_n$ satisfies a LDP with rate $n$ and rate function*

$$J(x) = \inf\{H(\nu|\mu) : \ \nu \in \mathcal{M}_{1,c}, \quad x = \rho'(\nu)\}. \tag{1.12}$$

*Here, $H(\nu|\mu)$ denotes the relative entropy of the probability measure $\nu$ with respect to $\mu$ defined by*

$$H(\nu|\mu) := \begin{cases} \int f \log f d\mu & \text{if } f := \frac{d\nu}{d\mu} \text{ exists} \\ \infty & \text{otherwise.} \end{cases} \tag{1.13}$$

*Proof.* The proof is a simple corollary of Sanov's Theorem (see e.g. Dembo and Zeitouni (1998), Theorem 6.2.10) and the contraction principle for empirical measures stated in Corollary 1.3.1. □

**Markov processes** A similar LDP for the risk measurements can be obtained if the observations are not independent, but are generated by certain stationary Markov chains. The empirical measures of Markov chains satisfying a strong uniformity condition fulfill a LDP. In this case we are able to derive a similar LDP for risk measurements as in the case of independent random variables. We remark that the results for Markov chains can also be extended to observations of stationary Markov processes in continuous time whose one-dimensional marginal distribution is equal to the law of the financial position $X$. Since this extension is technically more involved, we will stick to the discrete-time setup.

Let $(X_i)$ be a stationary Markov chain with state space $\mathbb{R}$ and one-dimensional marginal distribution $\mu$ where $\mu$ is the law of the financial position $X$ under the measure $P$. The transition probability measure will be denoted by $\pi(x, \cdot)$ $(x \in \mathbb{R})$. The $m$-step transition probability is defined recursively by

$$\pi^{m+1}(x, \cdot) = \int_{\mathbb{R}} \pi^m(y, \cdot)\pi(x, dy). \tag{1.14}$$

We need to make the following uniformity assumption which holds, in particular, for all finite state Markov chains which are irreducible. The condition goes back to Stroock (1984) and Ellis (1988).

***Assumption 1.3.3 (U).*** *There exist natural numbers* $0 < l \leq N$ *and a constant* $M \geq 1$ *such that for all* $x, y \in \mathbb{R}$

$$\pi^l(x, \cdot) \leq \frac{M}{N} \sum_{m=1}^{N} \pi^m(y, \cdot). \tag{1.15}$$

***Remark 1.3.4.*** *Note that the Assumption 1.3.3 (U) implies that the Markov chain is Doeblin recurrent.*

***Proposition 1.3.5.*** *Let* $\rho$ *be continuous on compacts, and suppose that assumption (U) holds. Then* $(\rho'(\mu_n))$ *satisfies a LDP with rate* $n$ *and rate function*

$$J(x) = \inf\{\Lambda^*(\nu) : \quad \nu \in \mathcal{M}_{1,c}, \quad x = \rho'(\nu)\}. \tag{1.16}$$

*Here,* $\Lambda^*$ *is the Fenchel-Legendre transform*

$$\Lambda^*(\nu) = \sup_{f \in C_b(\mathbb{R})} \left( \int f d\nu - \Lambda(f) \right) \tag{1.17}$$

*of the function*

$$\Lambda(f) = \lim_{n \to \infty} \frac{1}{n} \log E \left( \exp \left( \sum_{i=1}^{n} f(X_i) \right) \right) \tag{1.18}$$

*which is well-defined for every* $f \in C_b(\mathbb{R})$. $\Lambda^*$ *allows the alternative representation*

$$\Lambda^*(\nu) = \sup_{u \in C_b(\mathbb{R}),\ u \geq 1} \left\{ - \int_{\mathbb{R}} \log \left( \frac{\pi u}{u} \right) d\nu \right\} \tag{1.19}$$

*with*

$$\pi u := \int_{\mathbb{R}} u(y) \pi(\cdot, dy). \tag{1.20}$$

*Proof.* The proof is a direct consequence of the LDP for empirical measures (see Dembo and Zeitouni (1998), Exercise 6.3.13.) and the contraction theorem stated in Corollary 1.3.1. For the alternative representation of $\Lambda^*$ see Dembo and Zeitouni (1998), Section 6.5.1. □

Large deviations of risk measures for Markovian observations can also be obtained under different conditions than uniformity assumption (U). Let us assume that $(X_i)$ is an ergodic Markov process with one-dimensional marginal distribution $\mu$. We will assume that the process is hypermixing and need the following definition.

***Definition 1.3.6.*** *Let the natural numbers* $r \geq k \geq 2$, $l \geq 1$ *be given. A family* $(f_i)_{i=1}^k$ *of bounded, measurable functions on* $\mathbb{R}^r$ *is called* $l$-separated *if there exist* $k$ *disjoint intervals* $([a_i, b_i])_{i=1}^k$ *with* $a_i, b_i \in \{1, 2, 3, \ldots, r\}$, $a_i \leq b_i$ $(i = 1, 2, \ldots k)$ *and* $a_i - b_j \geq l$ *or* $a_j - b_i \geq l$ *for* $i \neq j$ *such that the function* $f_i(x_1, x_2, \ldots, x_r)$ *does actually depend only on the coordinates* $(x_j)_{j=a_i}^{b_i}$.

The following assumptions are taken from Dembo and Zeitouni (1998) and are referred to as hypermixing conditions. Hypermixing is related to analytical properties of the semigroup of the Markov chain and is closely linked to logarithmic Sobolev inequalities.

***Assumption 1.3.7 (H).*** *For a natural number $r$ we denote by $B(\mathbb{R}^r)$ the set of bounded, measurable functions on $\mathbb{R}^r$.*

*(1) There exist a natural numbers $l$ and a constant $\alpha \in (0, \infty)$ such that for all positive integers $k, r$ and any family of $l$-separated function $(f_i)_{i=1}^k \subseteq B(\mathbb{R}^r)$ the following inequality holds:*

$$E\left(\prod_{i=1}^{k} |f_i(X_1, \ldots X_r)|\right) \leq \prod_{i=1}^{k} E(|f_i(X_1, \ldots X_r)|^{\alpha})^{1/\alpha}. \tag{1.21}$$

*(2) There exists a natural number $l_0$ and functions $\beta(l) \geq 1$, $\gamma(l) \geq 0$ such that for all integers $l > l_0$, all natural numbers $r \geq 2$ and any two $l$-separated functions $f, g \in B(\mathbb{R}^r)$ the following conditions hold:*

*(a) $\lim_{l\to\infty} \gamma(l) = 0$ and $\limsup_{l\to\infty}(\beta(l) - 1)l(\log l)^{1+\delta}$ for some $\delta > 0$,*

*(b)*

$$\begin{aligned} &\Big| E[f(X_1, \ldots, X_r)] \cdot E[(g(X_1, \ldots, X_r)] \\ &\quad - \quad E[f(X_1, \ldots, X_r)(g(X_1, \ldots, X_r)] \Big| \\ &\leq \gamma(l) \cdot E\Big(|f(X_1, \ldots, X_r)|^{\beta(l)}\Big)^{1/\beta(l)} \; E\Big(|g(X_1, \ldots, X_r)|^{\beta(l)}\Big)^{1/\beta(l)} \end{aligned}$$

***Proposition 1.3.8.*** *Let $\rho$ be continuous on compacts, and suppose that assumption (H) holds. Then $(\rho'(\mu_n))$ satisfies a LDP with rate $n$ and rate function*

$$J(x) = \inf\{\Lambda^*(\nu) : \quad \nu \in \mathcal{M}_{1,c}, \quad x = \rho'(\nu)\}. \tag{1.22}$$

*Here, $\Lambda^*$ is the Fenchel-Legendre transform*

$$\Lambda^*(\nu) = \sup_{f \in B(\mathbb{R})} \left( \int f d\nu - \Lambda(f) \right) \tag{1.23}$$

*of the function*

$$\Lambda(f) = \lim_{n\to\infty} \frac{1}{n} \log E\left(\exp\left(\sum_{i=1}^{n} f(X_i)\right)\right) \tag{1.24}$$

*which is well-defined for every $f \in B(\mathbb{R})$. $\Lambda^*$ allows the alternative representation*

$$\Lambda^*(\nu) = \begin{cases} \sup_{u \in B(\mathbb{R}),\ u \geq 1} \left\{ -\int_{\mathbb{R}} \log\left(\frac{\pi u}{u}\right) d\nu \right\} & , \ \textit{if } \frac{d\nu}{d\mu} \textit{ exists} \\ \infty & , \ \textit{otherwise} \end{cases} \tag{1.25}$$

*with*

$$\pi u := \int_{\mathbb{R}} u(y)\pi(\cdot, dy). \tag{1.26}$$

*Proof.* The proof is a direct consequence of the LDP for empirical measures (see Dembo and Zeitouni (1998), p. 287 and p. 290) and the contraction theorem stated in Corollary 1.3.1. □

## 1.4 Entropy Minimization under $AVaR$-Constraints

As we have seen in the last section, for independently generated samples the rate function of the large deviations of risk measures is determined by the minimal relative entropy under a risk measure constraint. In the current section, we will discuss the minimization problem for a special risk measures: *average value at risk*. $AVaR$ is a risk measure with appealing properties. It is distribution-invariant and coherent. In the event of a large loss, $AVaR$ takes its size into account. The last fact follows, for example, from the following representation of average value at risk at level $\lambda$:

$$AVaR_\lambda(X) = \frac{1}{\lambda} E((q - X)^+) - q,$$

where $q$ is some $\lambda$-quantile of the random variable $X$.

**The minimization problem** Fix a reference probability measure $\mu \in \mathcal{M}_{1,c}$ with compact support, let $\lambda \in (0, 1)$ be a level and $y \in \mathbb{R}$ a constant. We are interested in the problem of minimizing $H(\nu|\mu)$ where $\nu \in \mathcal{M}_{1,c}$ and $AVaR_\lambda(\nu) = y$.

We set $a := \inf\{x \in \mathbb{R} : \ x \in \text{supp } \mu\}$, and $b := \sup\{x \in \mathbb{R} : \ x \in \text{supp } \mu\}$. Then supp $\mu \subseteq [a, b]$. Observe that supp $\nu \not\subseteq$ supp $\mu$ implies $\nu \not\ll \mu$, thus $H(\nu|\mu) = \infty$. We may therefore restrict our attention to the constraint set

$$\mathcal{C} := \{\nu \in \mathcal{M}_1([a, b]) : \ AVaR_\lambda(\nu) = y\}.$$

### 1.4.1 Existence of Solutions

A necessary and sufficient criterion for the existence of solutions can be formulated in terms of the parameters $a$, $b$ and $y$. We need the following general result that will also be used in Section 1.5. Since AVaR is continuous on compacts, the entropy minimization problem under a AVaR-constraint represents a special case of the next lemma.

Let $\mu \in \mathcal{M}_{1,c}$, and $a, b, y \in \mathbb{R}$ be given as above. For any distribution-invariant risk measure $\rho : L^\infty \to \mathbb{R}$ which is continuous on compacts we define

$$\mathcal{C}_\rho := \{\nu \in \mathcal{M}_1([a,b]) : \ \rho'(\nu) = y\}.$$

We consider the minimization problem of $H(\cdot|\mu)$ on $\mathcal{C}_\rho$.

***Lemma 1.4.1.*** *Suppose that $\mathcal{C}_\rho \neq \emptyset$. There exists a solution to the entropy minimization problem with constraint set $\mathcal{C}_\rho$. If there exists a $\nu \in \mathcal{C}_\rho$ such that $H(\nu|\mu) < \infty$, then the minimizer has finite relative entropy.*

*Proof.* If $H(\nu|\mu) = \infty$ for all $\nu \in \mathcal{C}_\rho$, then any $\nu \in \mathcal{C}_\rho$ minimizes the relative entropy. Otherwise observe that $\mathcal{M}_1([a,b])$ is weakly compact, since $[a,b]$ is compact. Since $\rho$ is continuous on compacts, $\mathcal{C}_\rho$ is a weakly compact set. Since $H(\cdot|\mu)$ is lower semicontinuous, it achieves its minimum on $\mathcal{C}_\rho$. $\square$

In the case of AVaR, the existence of a minimizer with finite relative entropy can be rephrased in terms of the parameters of the problem. For this purpose, it is useful to recall a particular representation of $AVaR_\lambda$, cf. Föllmer and Schied (2002c).

***Proposition 1.4.2.*** *Let $\lambda \in (0,1)$, and $\nu \in \mathcal{M}_{1,c}$. Then*

$$AVaR_\lambda(\nu) = -\int x f_\nu(x)\nu(dx), \tag{1.27}$$

*where $f_\nu$ is the following density of a probability measure with respect to $\nu$:*

$$f_\nu(x) = \frac{1}{\lambda}\left(\mathbf{1}_{(-\infty,q)} + \kappa \mathbf{1}_{\{q\}}\right) \tag{1.28}$$

*Here, $q$ is a $\lambda$-quantile of $\nu$, i.e.*

$$\int \mathbf{1}_{(-\infty,q)} d\nu \quad \leq \quad \lambda \tag{1.29}$$

$$\int \mathbf{1}_{(-\infty,q]} d\nu \quad \geq \quad \lambda \tag{1.30}$$

*The parameter $\kappa$ is defined as follows:*

$$\kappa = \begin{cases} 0 & \text{if } \nu\{q\} = 0 \\ \frac{\lambda - \nu(-\infty,q)}{\nu\{q\}} & \text{if } \nu\{q\} \neq 0 \end{cases} \tag{1.31}$$

The following proposition characterizes the existence of solutions.

***Proposition 1.4.3.*** *The following conditions are equivalent:*

*(1) There exists $\nu \in \mathcal{C}$ such that $H(\nu|\mu) < \infty$.*

*(2) The minimal value of the relative entropy on $\mathcal{C}$ is finite and attained for some element of $\mathcal{C}$.*

*(3) $a < -y < b$, or $-y$ is an atom of $\mu$.*

*Proof.* (1) and (2) are clearly equivalent by Lemma 1.4.1.

We will now show that (1) and (3) are equivalent. First suppose that $-y$ is an atom of $\mu$. Then we set

$$\frac{d\nu}{d\mu} = \frac{1}{\mu\{-y\}} \cdot \mathbf{1}_{\{-y\}}.$$

In this case, $-y$ is a $\lambda$-quantile of $\nu$, and $AVaR_\lambda(\nu) = y$.

Next suppose that $-y$ is not an atom of $\mu$ and that (1) holds. Then $-y \leq a$, $a < -y < b$, or $-y \geq b$. Let $\nu \in \mathcal{C}$ with $H(\nu|\mu) < \infty$. In particular, $\nu \ll \mu$. Since $\operatorname{supp} \nu \subseteq [a, b]$, it follows that $-y = -AVaR_\lambda(\nu) \in [a, b]$. If $a = -y = -AVaR_\lambda(\nu)$, then $a$ must be an atom of $\nu$. Since $\nu \ll \mu$, $a$ is then also an atom of $\mu$, a contradiction. Analogously, it can be shown that $b \neq -y$. We obtain therefore $a < -y < b$.

Finally, we have to show that for $a < -y < b$ there exists always $\nu \in \mathcal{C}$ such that $H(\nu|\mu) < \infty$. We consider two cases:

(a) There exists $q \in \mathbb{R}$ with $a < -y < q < b$ such that $\mu(-y, q) > 0$. Since $\mu$ has at most countably many atoms, we may and will assume that $q$ is not an atom of $\mu$.

(b) There exists no such $q \in \mathbb{R}$. This implies that $\mu(-y, b) = 0$. Then $b$ must be an atom of $\mu$, since $b = \sup\{x \in \mathbb{R} : x \in \operatorname{supp} \mu\}$.

We consider first case (a). Since $a < -y < q$, there exists $\alpha' \in (0, 1)$ such that

$$-y = \alpha' \frac{1}{\mu[a, -y)} \int x \mathbf{1}_{[a,-y)}(x)\mu(dx) + (1 - \alpha') \frac{1}{\mu[-y, q)} \int x \mathbf{1}_{[-y,q)}(x)\mu(dx). \quad (1.32)$$

Define the weights

$$\alpha = \frac{\lambda\alpha'}{\mu[a, -y)}, \quad \beta = \frac{\lambda(1 - \alpha')}{\mu[-y, q)}, \quad \gamma = \frac{1 - \lambda}{\mu[q, b]}. \quad (1.33)$$

We define a probability measure by

$$\frac{d\nu}{d\mu} = \alpha \mathbf{1}_{[a,-y)} + \beta \mathbf{1}_{[-y,q)} + \gamma \mathbf{1}_{[q,b]}. \quad (1.34)$$

Then $H(\nu|\mu) < \infty$. We show that $\nu \in \mathcal{C}$. First, by calculation we obtain that $\nu(-\infty, q) = \nu(-\infty, q] = \lambda$. Thus, $q$ is a $\lambda$-quantile of $\nu$. Second,

$$\begin{aligned}
AVaR_\lambda(\nu) &= -\frac{1}{\lambda}\int x\mathbf{1}_{[a,q)}(x)\nu(dx) \\
&= -\frac{1}{\lambda}\left(\alpha\int x\mathbf{1}_{[a,-y)}(x)\mu(dx) + \beta\int x\mathbf{1}_{[-y,q)}(x)\mu(dx)\right) \\
&= -\left(\frac{\alpha'}{\mu[a,-y)}\int x\mathbf{1}_{[a,-y)}(x)\mu(dx) + \frac{1-\alpha'}{\mu[-y,q)}\int x\mathbf{1}_{[-y,q)}(x)\mu(dx)\right) \\
&= y.
\end{aligned}$$

Next we consider case (b). In this case we set

$$\bar{a} := \frac{1}{\mu(-\infty,-y]}\int x\mathbf{1}_{(-\infty,-y]}\mu(dx) \; < \; -y.$$

Let $\gamma := \lambda(b+y)/(b-\bar{a}) \in (0,\lambda)$. We define a probability measure $\nu$ via its density

$$\frac{d\nu}{d\mu} = \frac{\gamma}{\mu[a,-y]}\mathbf{1}_{[a,b)} + \frac{1-\gamma}{\mu\{b\}}\mathbf{1}_{\{b\}}.$$

Then $H(\nu|\mu) < \infty$. We now verify that $\nu \in \mathcal{C}$. Observe that

$$\int \mathbf{1}_{(-\infty,b)}(x)\nu(dx) = \frac{\gamma}{\mu[a,-y]}\int \mathbf{1}_{(-\infty,-y]}(x)\mu(dx) = \gamma < \lambda.$$

This implies that $b$ is a $\lambda$-quantile for $\nu$. We can now calculate $AVaR_\lambda$ using (1.28). Here, $\kappa = (\lambda-\gamma)/\nu\{b\}$. Thus,

$$\begin{aligned}
AVaR_\lambda(\nu) &= \frac{-\gamma}{\lambda\cdot\mu[a,-y]}\int x\mathbf{1}_{(-\infty,b)}(x)\mu(dx) - \frac{(\lambda-\gamma)b}{\lambda} \\
&= \frac{-\bar{a}\gamma}{\lambda} - \frac{(\lambda-\gamma)b}{\lambda} \;=\; \frac{(b-\bar{a})\gamma}{\lambda} - b \;=\; y.
\end{aligned}$$

□

### 1.4.2 Structure of the Solutions

Classical results of Csiszar (1975) determine the general structure of the solution. We compute the solution explicitly. In order to avoid trivial cases, we will always assume that one and thus all of the equivalent conditions of Proposition 1.4.3 is satisfied. We distinguish two cases of different complexity.

(A) $\mu$ does not have any atoms.

(B) $\mu$ possibly has atoms.

**A Two-Step Procedure I**

First we focus on case (A). In the context of the entropy minimization problem, we can restrict our attention to probability measures which are absolutely continuous with respect to $\mu$. Since $\mu$ does not have any atoms, a minimizer $\nu$ will not have any atoms. Thus, if we calculate $AVaR_\lambda(\nu)$ according to (1.27), then the formulas characterizing the density $f_\nu$ simplify to

$$f_\nu = \frac{1}{\lambda}\mathbf{1}_{(-\infty,q)},$$
$$\int \mathbf{1}_{(-\infty,q)}(x)\nu(dx) = \lambda.$$

The original problem can be reduced to a family of relative entropy minimization problems under linear constraints and a one-dimensional minimization problem.

**Step 1** Fix some quantile level $q \in \mathbb{R}$. Minimize $\nu \mapsto H(\nu|\mu)$ over all probability measures $\nu \ll \mu$ which satisfy the constraint

$$-\frac{1}{\lambda}\int x\mathbf{1}_{(-\infty,q)}(x)\nu(dx) = y, \tag{1.35}$$
$$\int \mathbf{1}_{(-\infty,q)}(x)\nu(dx) = \lambda. \tag{1.36}$$

We will provide conditions under which this problem has a solution. Then the solution is unique and can be represented by an exponential density.

**Step 2** As we will see, if for $q \in \mathbb{R}$ the minimization problem in step 1 has a solution with finite relative entropy, the minimizer will be unique. We denote this minimizer by $\nu^q$. Otherwise, we set $\nu^q = \dagger$ with the convention $H(\dagger|\mu) = \infty$. With this notation, the solution of the original problem is given by the set

$$\mathrm{argmin}_{\nu\in\mathcal{D}}\, H(\nu|\mu), \quad \mathcal{D} = \{\nu^q :\ q \in \mathbb{R}\}.$$

Since the original problem has a solution with finite relative entropy under the conditions specified in Proposition 1.4.3, $\mathcal{D}$ contains at least one element different from $\dagger$. Thus, for some $q \in \mathbb{R}$ the minimization problem in step 1 will have a solution with finite relative entropy. We will now discuss the two steps of the solution strategy in more detail.

**Entropy Minimization under Linear Constraints I**

We fix an arbitrary reference measure $\mu \in \mathcal{M}_{1,c}$ and $q \in \mathbb{R}$. We assume that the measure $\mu$ does not have any atoms. In particular, $q$ is *not* an atom of $\mu$. In this section we consider the following minimization problem: minimize $\nu \mapsto H(\nu|\mu)$ over all probability measures which satisfy the constraint (1.35) and (1.36).

***Proposition 1.4.4.*** *The following conditions are equivalent:*

*(1) There exists a probability measure $\nu$ with $H(\nu|\mu) < \infty$ that satisfies the constraint.*

*(2) There exists a probability measure $\nu$ equivalent to $\mu$ with $H(\nu|\mu) < \infty$ that satisfies the constraint.*

*(3) Under the constraint there exists a unique minimizer of the relative entropy.*

*(4) $a < -y < q < b$, $\mu(-y, q) > 0$.*

*Proof.* (3) trivially implies (1). In order to show that (1) implies (3) observe that the constraint set defined by (1.35) and (1.36) is variation-closed and convex. Theorem 2.1. of Csiszar (1975) implies that the minimization problem has a solution with finite relative entropy. The uniqueness of the minimizer follows, since the constraint set is convex and $H(\cdot|\mu)$ is strictly convex on its essential domain. Altogether, we have shown that (1) and (3) are equivalent.

Next, we show that (2) $\Rightarrow$ (1) $\Rightarrow$ (4) $\Rightarrow$ (2). The first implication is clear. Assume that (1) holds. We show that this implies (4). By assumption, $\nu \ll \mu$, and $\nu$ does not have any atoms. Since supp $\nu \subseteq [a, b]$, we obtain that $q \in (a, b)$. Thus,

$$-y \;=\; \frac{1}{\lambda}\int x\mathbf{1}_{(-\infty,q)}(x)\nu(dx) \;=\; \frac{1}{\nu[a,q)}\int x\mathbf{1}_{[a,q)}(x)\nu(dx) \;\in\; [a,q). \tag{1.37}$$

$-y$ is not an atom of $\mu$. Then (1.37) implies that $a < -y$. Suppose moreover $\mu(-y, q) = 0$, thus $\nu[-y, q) = 0$. Then,

$$-y = \frac{1}{\lambda}\int x\mathbf{1}_{(-\infty,q)}(x)\nu(dx) = \frac{1}{\nu(a,-y)}\int x\mathbf{1}_{(a,-y)}(x)\nu(dx) < -y,$$

a contradiction.

Finally, we show that (4) implies (2). Define the density of $\nu$ with respect to $\mu$ by (1.34) with coefficients given by (1.32) and (1.33). This defines a measure $\nu$ which is equivalent to $\mu$ and satisfies the constraints (1.35) and (1.36). □

If one and thus all of the equivalent conditions of Proposition 1.4.4 are satisfied, the unique minimizer can be characterized. Its density with respect to $\mu$ is of exponential form. The exponent is a linear combination of the constraint functions. We quote a theorem of Csiszar (1975).

**Theorem 1.4.5.** *For $i = 1, 2, \ldots, I$ let $f_i : \mathbb{R} \to \mathbb{R}$ be measurable functions and $a_i \in \mathbb{R}$. Let $\mu \in \mathcal{M}_1(\mathbb{R})$, and define the constraint set*

$$\hat{\mathcal{C}} = \left\{ \nu \in \mathcal{M}_1(\mathbb{R}) : \int f_i(x)\nu(dx) = a_i, \ \ i = 1, 2, \ldots, I \right\}.$$

*Assume there exists $\nu \in \hat{\mathcal{C}}$ with $\nu \approx \mu$ and $H(\nu|\mu) < \infty$. Then there exists a unique minimizer on $\hat{\mathcal{C}}$ with finite relative entropy. $\nu$ is the minimizer, if and only if its $\mu$-density is of the following form*

$$\frac{d\nu}{d\mu} \ = \ c \cdot \exp\left( \sum_{i=1}^{I} h_i f_i \right),$$

*with normalizing constant $c > 0$ and $h_i \in \mathbb{R}$ $(i = 1, 2, \ldots, I)$.*

**Corollary 1.4.6.** *Assume that one of the equivalent conditions of Proposition 1.4.4 holds. $\nu$ is the unique minimizer of the relative entropy under the constraints (1.35) and (1.36), if and only if its $\mu$-density is of the following form:*

$$\frac{d\nu}{d\mu}(x) = c \cdot \exp\left( (h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x) \right). \tag{1.38}$$

*Here, $c > 0$ is a normalizing constant, $h_1, h_2 \in \mathbb{R}$, and the following conditions need to be satisfied:*

$$-\lambda y \ = \ c \cdot \int x \exp(h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x)\mu(dx), \tag{1.39}$$

$$\lambda \ = \ c \cdot \int \exp(h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x)\mu(dx), \tag{1.40}$$

$$1 \ = \ c \cdot \int \exp\left\{ (h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x) \right\} \mu(dx). \tag{1.41}$$

*Proof.* The proof follows directly from Theorem 1.4.5. (1.41) is a normalization, (1.39) and (1.40) are required by the constraint. □

If $\nu$ is the minimizing density characterized in Corollary 1.4.6, the minimal relative entropy is given by the expression

$$\begin{aligned} H(\nu|\mu) \ &= \ c \log c \int \exp\left( (h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x) \right) \mu(dx) \\ &\quad + \ c \int (h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x) \exp(h_1 + h_2 x)\mu(dx) \\ &= \ \log c \ + \ \lambda(h_1 - h_2 y). \end{aligned}$$

### The Original Problem I

Assuming that one and thus all of the equivalent conditions of Proposition 1.4.4 are satisfied, we denote the unique solution of the minimization problem under the linear constraints (1.35) and (1.36) by $\nu^q$.

***Proposition 1.4.7.*** *There exists a solution to the minimization problem of the relative entropy on $\mathcal{C}$, if and only if $a < -y < b$. The collection of all solutions is given by $\{\nu^q : q \in Q^*\}$ where $Q^* =$ argmin $\{H(\nu^q|\mu) : a < -y < q < b, \ \mu(-y, q) > 0\}$.*

*Proof.* Since $\mu$ does not have any atoms, according to Proposition 1.4.3, a solution with finite relative entropy exists, if and only if $a < -y < b$. Clearly, all solutions are obtained by solving first the minimization problem under constraints (1.35) and (1.36) for fixed $q \in \mathbb{R}$ and then by minimizing over those $q \in \mathbb{R}$ for which a solution with finite entropy exists. □

### A Two-Step Procedure II

We now consider case (B), i.e. $\mu$ may have atoms. As in case (A) the problem can be decomposed into two subproblems, but atoms make the problem more complicated.

**Step 1** Fix some quantile level $q \in \mathbb{R}$. We distinguish two cases:

a) $q$ is not an atom of the reference measure $\mu$.

b) $q$ is an atom of $\mu$.

Case a) is only slightly more complicated than the situation which we considered in Proposition 1.4.4. The first step is to minimize $\nu \mapsto H(\nu|\mu)$ over all probability measures $\nu \ll \mu$ which satisfy the constraints (1.35) and (1.36).

Case b) involves two additional parameters. Let $d \in [0, \lambda]$ and $u \in [0, 1-\lambda]$. If $d = u = 0$, we set $d \cdot (d+u)^{-1} = 0$. In step 1 we need to minimize $\nu \mapsto H(\nu|\mu)$ over all probability measures $\nu \ll \mu$ which satisfy the constraint

$$\lambda - d = \int \mathbf{1}_{(-\infty,q)}(x)\nu(dx), \tag{1.42}$$

$$\lambda + u = \int \mathbf{1}_{(-\infty,q]}(x)\nu(dx), \tag{1.43}$$

$$y = -\frac{1}{\lambda}\int x\left(\mathbf{1}_{(-\infty,q)}(x) + \frac{d}{d+u}\cdot \mathbf{1}_{\{q\}}(x)\right)\nu(dx). \tag{1.44}$$

In both cases, we will provide conditions when the problems have a solution. Then the solution is unique and can again be represented by a density which is of exponential form outside the set where it vanishes. The solution will not always be equivalent to the reference measure $\mu$.

**Step 2** As we will show, if the minimization problems a) and b) in step 1 have a solution with finite relative entropy for fixed parameters $q$, $u$, $d$, then the minimizer will be unique. Let $q \in \mathbb{R}$, $d \in [0, \lambda]$ and $u \in [0, 1-\lambda]$. If $q$ is not an atom of $\mu$, we are in the situation of case a). If the minimizer exists and if $u = d = 0$, we denote it by $\nu^{q,d,u} = \nu^{q,0,0}$. If $q$ is not an atom, we consider case b). If the minimizer exists, we denote it by $\nu^{q,d,u}$. In all other cases, we set $\nu^{q,u,d} = \dagger$ with the convention $H(\dagger|\mu) = \infty$. With this notation, the solution of the original problem is given by the set of minimizers

$$\operatorname{argmin}_{\nu \in \mathcal{D}} H(\nu|\mu), \quad \mathcal{D} = \{\nu^{q,d,u} : \ q \in \mathbb{R},\ d \in [0, \lambda],\ u \in [0, 1-\lambda]\}. \tag{1.45}$$

Since the original problem has a solution with finite relative entropy under the conditions specified in Proposition 1.4.3, $\mathcal{D}$ contains at least one element different from $\dagger$. Thus, for some triple $q \in \mathbb{R}$, $d \in [0, \lambda]$, $u \in [0, 1-\lambda]$ the minimization problem in step 1 will have a solution with finite relative entropy. We will now discuss the two steps of the solution strategy in more detail.

### Entropy Minimization under Linear Constraints II

**Case a)** We fix an arbitrary reference measure $\mu \in \mathcal{M}_{1,c}$ and $q \in \mathbb{R}$. We assume that $q$ is *not* an atom of $\mu$. Nevertheless, the measure $\mu$ may have atoms. In this section we consider the following minimization problem: minimize $\nu \mapsto H(\nu|\mu)$ over all probability measures which satisfy the constraint (1.35) and (1.36). This problem is closely related to the minimization problem in Section 1.4.2. Nevertheless, if $\mu$ does have atoms, the situation is slightly more complicated: in certain cases the minimizer will not anymore be equivalent to the reference measure $\mu$.

***Proposition 1.4.8.*** *The following conditions are equivalent:*

*(1) There exists a probability measure $\nu$ with $H(\nu|\mu) < \infty$ that satisfies the constraint.*

*(2) Under the constraint there exists a unique minimizer of the relative entropy.*

*(3) One of the following conditions holds:*

*(a) $a < -y < q < b$, $\mu(-y, q) > 0$.*

*(b) $a \leq -y < q < b$, $-y$ is an atom of $\mu$.*

*Moreover, if condition (3)(a) holds, then there exists a probability measure $\nu$ equivalent to $\mu$ with $H(\nu|\mu) < \infty$ that satisfies the constraint.*

*Proof.* Proving that (1) and (2) are equivalent, is completely analogous to the proof of the equivalence of (1) and (3) in Proposition 1.4.4.

Next, we show that (1) and (3) are equivalent. Assume that (1) holds. By assumption, $\nu \ll \mu$, and $\nu$ does not have any atom at $q$. Since $\operatorname{supp} \nu \subseteq [a, b]$, we obtain that $q \in (a, b)$. Thus,

$$-y \;=\; \frac{1}{\lambda} \int x \mathbf{1}_{(-\infty,q)}(x)\nu(dx) \;=\; \frac{1}{\nu[a,q)} \int x \mathbf{1}_{[a,q)}(x)\nu(dx) \;\in\; [a,q). \tag{1.46}$$

Suppose that $-y$ is not an atom of $\mu$. Then (1.46) implies that $a < -y$. Suppose moreover $\mu(-y, q) = 0$, thus $\nu[-y, q) = 0$. Then,

$$-y = \frac{1}{\lambda} \int x \mathbf{1}_{(-\infty,q)}(x)\nu(dx) = \frac{1}{\nu[a,-y)} \int x \mathbf{1}_{[a,-y)}(x)\nu(dx) < -y,$$

a contradiction.

Finally, we show that (3) implies (1). If $-y$ is an atom for $\mu$, we set

$$\frac{d\nu}{d\mu} \;=\; \frac{\lambda}{\mu\{-y\}} \cdot \mathbf{1}_{\{-y\}} \;+\; \frac{1-\lambda}{\mu(q,b]} \cdot \mathbf{1}_{(q,b]}. \tag{1.47}$$

The measure $\nu$ satisfies the constraints (1.35) and (1.36) and has finite relative entropy. Nevertheless, it might not be equivalent to $\mu$.

Otherwise, $a < -y$ and $\mu(-y, q) > 0$. Define the density of $\nu$ with respect to $\mu$ by (1.34) with coefficients given by (1.32) and (1.33). This defines a measure $\nu$ which is equivalent to $\mu$ and satisfies the constraints (1.35) and (1.36). □

***Definition 1.4.9.*** *For $i = 1, 2, \ldots, I$ let $f_i : \mathbb{R} \to \mathbb{R}$ be measurable functions and $a_i \in \mathbb{R}$. Let $\mu \in \mathcal{M}_1(\mathbb{R})$, and define the constraint set*

$$\hat{\mathcal{C}} = \left\{ \nu \in \mathcal{M}_1(\mathbb{R}) : \int f_i(x)\nu(dx) = a_i, \;\; i = 1, 2, \ldots, I \right\}.$$

*By $\bar{\mathcal{C}}$ we denote the subset of elements $\nu \in \hat{\mathcal{C}}$ with $H(\nu|\mu) < \infty$. A measurable set $N \subseteq \mathbb{R}$ is called* maximal common nullset *if*

*(1) $\forall\, \nu \in \bar{\mathcal{C}}$: $\nu(N) = 0$,*

*(2) $\exists\, \nu \in \bar{\mathcal{C}}$: $\mathbf{1}_N + \frac{d\nu}{d\mu} > 0$ $\mu$-almost surely.*

**Remark 1.4.10.** *A maximal common nullset is indeed a maximal set in the following sense: Let $M \subseteq \mathbb{R}$ be a measurable set that satisfies condition (1) of Definition 1.4.9, and assume that $N \subseteq M$ is a maximal common nullset. Then $M$ is also a maximal common nullset, and $\mu(M \setminus N) = 0$ .*

*Proof.* It is clear that $M$ is a maximal common nullset. Let $\nu \in \bar{\mathcal{C}}$ satisfy property (2) of Definition 1.4.9 for the maximal common nullset $N$. Since $\nu(M) = 0$, we have $\frac{d\nu}{d\mu} \cdot \mathbf{1}_M = 0$ $\mu$-almost surely. Thus, $\mu$ almost surely

$$\left(\mathbf{1}_N \ + \ \frac{d\nu}{d\mu}\right) \cdot \mathbf{1}_{M\setminus N} \ = \ 0.$$

Observing that $\mathbf{1}_N + \frac{d\nu}{d\mu} > 0$ $\mu$-almost surely, this implies $\mu(M \setminus N) = 0$. □

The following remark is elementary, but clarifies under which conditions $\mu$-equivalent elements of $\bar{\mathcal{C}}$ exist.

**Remark 1.4.11.** *The following conditions are equivalent:*

*(1) There exists $\nu \in \bar{\mathcal{C}}$ with $\nu \approx \mu$.*

*(2) Some maximal common nullset is a $\mu$-nullset.*

*(3) Any maximal common nullset is a $\mu$-nullset.*

*(4) Any $\mu$-nullset is a maximal common nullset.*

*(5) Maximal common nullset and $\mu$-nullsets coincide.*

*(6) The empty set is a maximal common nullset.*

*Proof.*
(1) ⇒ (3): If $N$ is a maximal common nullset, then $\nu(N) = 0$. Thus $\mu(N) = 0$, since $\mu \ll \nu$.
(3) ⇒ (2): This is obvious.
(2) ⇒ (4): Let $N$ be a maximal common nullset with $\mu(N) = 0$, and let $M$ be a $\mu$-nullset. Let $\nu \in \bar{\mathcal{C}}$. Then $\nu \ll \mu$, thus $\nu(M) = 0$.

Choose $\nu \in \bar{\mathcal{C}}$ such that $1_N + \frac{d\nu}{d\mu} > 0$ $\mu$-almost surely. Since $N$ is a $\mu$-nullset, we obtain that $\frac{d\nu}{d\mu} > 0$ $\mu$-almost surely, thus $1_M + \frac{d\nu}{d\mu} > 0$ $\mu$-almost surely.
(4) ⇒ (6): This is obvious.
(6) ⇒ (1): Since the empty set is a maximal common nullset, there exists $\nu \in \bar{\mathcal{C}}$ with $\frac{d\nu}{d\mu} > 0$ $\mu$-almost surely. Thus, $\nu \gg \mu$. Conversely, $\mu \gg \nu$ by $H(\nu|\mu) < \infty$.
(5) ⇔ (3): Finally, observe that (5) is equivalent to the other conditions, since (3) and (4) are equivalent. □

In the context of the minimization problem of the current section maximal common nullsets can be characterized in terms of the parameters of the problem. If condition (3)(a) of Proposition 1.4.8 holds, maximal common nullset are $\mu$-nullsets. The next proposition investigates maximal common nullsets, if condition (3)(a) is *not* satisfied, but condition (3)(b) holds.

**Proposition 1.4.12.** *Assume that condition (3)(b) of Proposition 1.4.8 holds.*

*(1) If $a = -y$ and $\mu(-y, q) = 0$, then any maximal common nullset is a $\mu$-nullset. I.e. there exists a $\mu$-equivalent probability measure $\nu$ with $H(\nu|\mu) < \infty$ that satisfies the constraint.*

*(2) If $a = -y$ and $\mu(-y, q) > 0$ , then $(a, q)$ is a maximal common nullset.*

*(3) If $a < -y$ and $\mu(-y, q) = 0$, then $[a, -y)$ is a maximal common nullset.*

*Proof.* In case (1) equation (1.47) defines a density of a $\mu$-equivalent probability measure $\nu$ with $H(\nu|\mu) < \infty$ that satisfies the constraint. In order to verify (2), set $N := (a, q)$. Let $\nu$ be a measure with $H(\nu|\mu) < \infty$ that satisfies the constraint. If $\nu(N) > 0$, then

$$-y = \frac{1}{\lambda} \int x \mathbf{1}_{(-\infty,q)} \nu(dx) > -y, \tag{1.48}$$

a contradiction. Next, define a measure $\nu$ by density (1.47). As shown in the proof of Proposition 1.4.8, $\nu$ satisfies the constraints (1.35) and (1.36), and $H(\nu|\mu) < \infty$. Moreover, $\mu$-almost surely,

$$\mathbf{1}_N \; + \; \frac{d\nu}{d\mu} \; > \; 0.$$

The proof of (3) is completely analogous to the proof of (2). We simply have to set $N := [a, -y)$ and to reverse the inequality in (1.48). $\square$

The minimizers are characterized by the following theorem of Csiszar (1975).

**Theorem 1.4.13.** *For $i = 1, 2, \dots, I$ let $f_i : \mathbb{R} \to \mathbb{R}$ be measurable functions and $a_i \in \mathbb{R}$. Let $\mu \in \mathcal{M}_1(\mathbb{R})$, and define the constraint set*

$$\hat{\mathcal{C}} = \left\{ \nu \in \mathcal{M}_1(\mathbb{R}) : \int f_i(x)\nu(dx) = a_i, \;\; i = 1, 2, \dots, I \right\}.$$

*Assume there exists $\nu \in \hat{\mathcal{C}}$ with $H(\nu|\mu) < \infty$. Let $N$ be a maximal common nullset. Then there exists a unique minimizer on $\hat{\mathcal{C}}$ with finite relative entropy. $\nu$ is the minimizer, if and only if its $\mu$-density is of the following form*

$$\frac{d\nu}{d\mu} \; = \; c \cdot \exp\left( \sum_{i=1}^{I} h_i f_i \right) \cdot \mathbf{1}_{N^c},$$

*with normalizing constant $c > 0$ and $h_i \in \mathbb{R}$ $(i = 1, 2, \dots, I)$.*

***Corollary 1.4.14.*** *Assume that one and thus all of the equivalent conditions of Proposition 1.4.8 hold. Let $N$ be a maximal common nullset, cf. Propositions 1.4.8 & 1.4.12. $\nu$ is the unique minimizer of the relative entropy under the constraints (1.35) and (1.36), if and only if its $\mu$-density is of the following form:*

$$\frac{d\nu}{d\mu}(x) = c \cdot \exp\left((h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x)\right) \cdot \mathbf{1}_{N^c}(x). \tag{1.49}$$

*Here, $c > 0$ is a normalizing constant, $h_1, h_2 \in \mathbb{R}$, and the following conditions need to be satisfied:*

$$-\lambda y = c \cdot \int x \exp(h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x) \cdot \mathbf{1}_{N^c}(x)\mu(dx), \tag{1.50}$$

$$\lambda = c \cdot \int \exp(h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x) \cdot \mathbf{1}_{N^c}(x)\mu(dx), \tag{1.51}$$

$$1 = c \cdot \int \exp\left\{(h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x)\right\} \cdot \mathbf{1}_{N^c}(x)\mu(dx). \tag{1.52}$$

*Proof.* The proof follows directly from Theorem 1.4.13. (1.52) is a normalization, (1.50) and (1.51) are required by the constraint. □

If $\nu$ is the minimizing density characterized in Corollary 1.4.14, the minimal relative entropy is given by the expression

$$\begin{aligned} H(\nu|\mu) &= c \log c \int \exp\left((h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x)\right) \cdot \mathbf{1}_{N^c}(x)\, \mu(dx) \\ &\quad + c \int (h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x) \exp(h_1 + h_2 x) \cdot \mathbf{1}_{N^c}(x)\, \mu(dx) \\ &= \log c + \lambda(h_1 - h_2 y). \end{aligned}$$

**Case b)** We fix an arbitrary reference measure $\mu \in \mathcal{M}_{1,c}$ and parameters $q \in \mathbb{R}$, $d \in [0, \lambda]$ and $u \in [0, 1 - \lambda]$. Now we assume that $q$ *is* an atom of $\mu$. In this section we consider the following minimization problem: minimize $\nu \mapsto H(\nu|\mu)$ over all probability measures which satisfy the constraint (1.42), (1.43) and (1.44).

***Proposition 1.4.15.*** *The following conditions are equivalent:*

*(1) There exists a probability measure $\nu$ with $H(\nu|\mu) < \infty$ that satisfies the constraint.*

*(2) Under the constraint there exists a unique minimizer of the relative entropy.*

*(3) $a \le -y \le q \le b$, and one of the following conditions holds:*

*(a) $d = 0$, $-y < q$, and $-y$ is an atom of $\mu$*

*(b)* $d = 0$, $a < -y < q$, $\mu(-y, q) > 0$

*(c)* $d > 0$, $a < -y$ *and with* $\bar{a} := \sup\{x \in \mathit{supp}\ \mu : \ x < q\}$

$$-y \quad > \quad \frac{\lambda - d}{\lambda} \cdot a \ + \ \frac{d}{\lambda} \cdot q$$

$$-y \quad < \quad \frac{\lambda - d}{\lambda} \cdot \bar{a} \ + \ \frac{d}{\lambda} \cdot q$$

*(d)* $d > 0$, *and for some atom* $r \in \mathbb{R}$ *of* $\mu$,

$$-y \quad = \quad \frac{\lambda - d}{\lambda} \cdot r \ + \ \frac{d}{\lambda} \cdot q$$

*Moreover, if conditions (3)(b) holds, or if condition (3)(c) holds, then there exists a probability measure* $\nu$ *equivalent to* $\mu$ *with* $H(\nu|\mu) < \infty$ *that satisfies the constraint.*

*Proof.* Proving that (1) and (2) are equivalent, is analogous to the proof of the equivalence of (1) and (3) in Proposition 1.4.4.

We now show that (1) implies (3). First consider the case $d = 0$. Then

$$-y = \frac{1}{\lambda} \int x \mathbf{1}_{(-\infty,q)}(x) \nu(dx) = \frac{1}{\nu(-\infty, q)} \int x \mathbf{1}_{(-\infty,q)}(x) \nu(dx) < q. \tag{1.53}$$

If $-y$ is an atom of $\mu$, then (a) holds. Next, suppose that $-y$ is not an atom of $\mu$. Then it is not an atom of $\nu$. From (1.53) follows that $-y > a$. Suppose $\mu(-y, q) = 0$. Then $\nu(-y, q) = 0$. Thus, $\nu(-\infty, -y) = \nu(-\infty, q) = \lambda$. Hence,

$$-y = \frac{1}{\lambda} \int x \mathbf{1}_{(-\infty,q)}(x) \nu(dx) = \frac{1}{\nu(-\infty, -y)} \int x \mathbf{1}_{(-\infty,-y)}(x) \nu(dx) < -y,$$

a contradiction. This implies (b).

Next consider the case $d > 0$. Then $q$ is an atom for $\nu$. If $\lambda = d$, then (d) holds with $r = q$. Otherwise, $\nu(-\infty, q) > 0$ and $a < q$. We obtain from (1.44),

$$-y \quad = \quad \frac{\lambda - d}{\lambda} \underbrace{\frac{1}{\nu(-\infty, q)} \int x \mathbf{1}_{(-\infty,q)}(x) \nu(dx)}_{=:r} \ + \ \frac{d}{\lambda} \cdot q.$$

Then $r \in [a, q)$ and $a < -y$. If $r$ is not an atom of $\mu$, it is not an atom of $\nu$ and $a < r < \bar{a}$. This implies (c). If $r$ is an atom of $\mu$, then (d) holds.

Finally, we have to show that (3) implies (1). In all cases we will specify a density with respect to $\mu$ such that the resulting measure $\nu$ satisfies $H(\nu|\mu) < \infty$ and the constraint. In case (a) choose

$$\frac{d\nu}{d\mu} = \frac{\lambda}{\mu\{-y\}} \cdot \mathbf{1}_{\{-y\}} + \frac{1 - \lambda}{\nu[q, \infty)} \cdot \mathbf{1}_{[q,\infty)}. \tag{1.54}$$

Next, consider case (b). Then clearly

$$\frac{1}{\mu[-y,q)}\int x\mathbf{1}_{[-y,q)}(x)\mu(dx) > -y.$$

Hence, there exists $\alpha' \in (0,1)$ which satisfies (1.32). We choose a density according to (1.33) and (1.34). As in the proof of Proposition 1.4.3 simple calculations show that the constraints are satisfied. Observe that the measure $\nu$ specified by (1.34) is equivalent to $\mu$.

Assume that (3)(c) is satisfied. Then there exists $r \in (a, \bar{a})$ such that

$$-y = \frac{\lambda - d}{\lambda}\cdot r + \frac{d}{\lambda}\cdot q.$$

By the definition of $a$ and $\bar{a}$ it holds that $\mu[a,r) > 0$ and $\mu[r,q) > 0$. Moreover,

$$\begin{aligned} r &> \frac{1}{\mu[a,r)}\int x\mathbf{1}_{[a,r)}(x)\mu(dx) =: g_-, \\ r &< \frac{1}{\mu[r,q)}\int x\mathbf{1}_{[r,q)}(x)\mu(dx) =: g_+. \end{aligned}$$

Thus, there exists $\alpha \in (0,1)$ such that $r = \alpha g_- + (1-\alpha)g_+$. Define

$$\frac{d\nu}{d\mu} = (\lambda - d)\left(\frac{\alpha}{\mu[a,r)}\cdot\mathbf{1}_{[a,r)} + \frac{1-\alpha}{\mu[r,q)}\cdot\mathbf{1}_{[r,q)}\right) + \frac{d+u}{\mu\{q\}}\cdot\mathbf{1}_{\{q\}} + \frac{1-\lambda-u}{\mu(q,\infty)}\cdot\mathbf{1}_{(q,\infty)}.$$

Then $\nu$ satisfies the constraints (1.42), (1.43) and (1.44). Observe that $\nu$ is equivalent to $\mu$.

Finally, if (3)(d) holds, then e.g. the following density defines an appropriate measure $\mu$:

$$\frac{d\nu}{d\mu} = \frac{\lambda - d}{\mu\{r\}}\cdot\mathbf{1}_{\{r\}} + \frac{d+u}{\mu\{q\}}\cdot 1_{\{q\}} + \frac{1-\lambda-u}{\mu(q,\infty)}\cdot\mathbf{1}_{(q,\infty)}. \tag{1.55}$$

□

The next proposition investigates maximal common nullsets, if the conditions in part (3) of Proposition 1.4.15 are satisfied. This characterization together with Theorem 1.4.13 will allow us to specify the solution of the minimization problem of the relative entropy $\nu \mapsto H(\nu|\mu)$ under the constraint (1.42), (1.43) and (1.44).

***Proposition 1.4.16.*** *Assume that $a \le -y \le q \le b$ holds.*

*(1) Suppose that $d = 0$ and that $-y = a < q$ is an atom of $\mu$. Then $(-y,q)$ is a maximal common nullset.*

*(2) Suppose $d = 0$ and $a < -y < q$. If $\mu(-y, q) = 0$ and $-y$ is an atom of $\mu$, then $[a, -y)$ is a maximal common nullset. If $\mu(-y, q) > 0$, then the empty set is a maximal common nullset.*

*(3) If condition (3)(c) of Proposition 1.4.15 holds, the empty set is a maximal common nullset.*

*(4) Suppose that condition (3)(d) of Proposition 1.4.15 holds.*

*(a) If $\lambda = d$, then $(-\infty, q)$ is a maximal common nullset.*

*(b) If $\lambda \neq d$ and $r = a$, then $(a, q)$ is a maximal common nullset.*

*(c) If $\lambda \neq d$, $r > a$ and $\mu(r, q) = 0$, then $[a, r)$ is a maximal common nullset.*

*Proof.* Denote by $\bar{\mathcal{C}}$ the set of measures $\nu$ with $H(\nu|\mu) < \infty$ that satisfies the constraints (1.42),(1.43) and(1.44). It follows from Proposition 1.4.15 that $\bar{\mathcal{C}}$ is the never empty for the cases considered in the current proposition.
ad (1): Let $\nu \in \bar{\mathcal{C}}$. If $\nu(-y, q) > 0$, then (1.44) implies that

$$-y = \frac{\nu\{-y\}}{\nu[-y, q)} \cdot (-y) + \frac{\nu(-y, q)}{\nu[-y, q)} \cdot \frac{1}{\nu(-y, q)} \int x\mathbf{1}_{(-y,q)}(x)\nu(dx) > -y,$$

a contradiction. Thus, $\nu(-y, q) = 0$.

Next, define $\nu \in \bar{\mathcal{C}}$ by density (1.54). Then, $\mu$-almost surely $\mathbf{1}_{(-y,q)} + \frac{d\nu}{d\mu} > 0$.
ad (2): We consider first the case $\mu(-y, q) = 0$. Let $\nu \in \bar{\mathcal{C}}$. Assume that $\nu[a, -y) > 0$. Then (1.44) implies that

$$-y = \frac{\nu\{-y\}}{\nu[a, -y]} \cdot (-y) + \frac{\nu[a, -y)}{\nu[a, -y]} \cdot \frac{1}{\nu[a, -y)} \int x\mathbf{1}_{[a,-y)}(x)\nu(dx) < -y,$$

a contradiction. Thus, $\nu[a, -y) = 0$.

If $\nu \in \bar{\mathcal{C}}$ is specified via density (1.54), then $\mu$-almost surely $\mathbf{1}_{[a,-y)} + \frac{d\nu}{d\mu} > 0$.

Secondly, we consider the case $\mu(-y, q) > 0$. By Proposition 1.4.15 there exists a $\mu$ equivalent $\nu \in \bar{\mathcal{C}}$. The claim follows from Remark 1.4.11.
ad (3): This follows from Proposition 1.4.15 and Remark 1.4.11.
ad (4): If $\lambda = d$, then $\nu(-\infty, q) = 0$ for $\nu \in \bar{\mathcal{C}}$ by (1.42).

In case (b) assume for $\nu \in \bar{\mathcal{C}}$ that $\nu(a, q) > 0$. Then we obtain from (1.44),

$$-y = \frac{1}{\lambda} \int x\mathbf{1}_{(-\infty,q)}(x)\nu(dx) + \frac{d}{\lambda} \cdot q = \frac{\lambda - d}{\lambda} \cdot \frac{1}{\lambda - d} \int x\mathbf{1}_{[r,q)}(x)\nu(dx) + \frac{d}{\lambda} \cdot q > -y,$$

a contradiction. Thus, $\nu(a, q) = 0$.

In case (c) assume for $\nu \in \bar{\mathcal{C}}$ that $\nu[a, r) > 0$. Then we obtain from (1.44),

$$-y = \frac{1}{\lambda} \int x \mathbf{1}_{(-\infty,q)}(x)\nu(dx) + \frac{d}{\lambda} \cdot q = \frac{\lambda - d}{\lambda} \cdot \frac{1}{\lambda - d} \int x \mathbf{1}_{[a,r)}(x)\nu(dx) + \frac{d}{\lambda} \cdot q < -y,$$

a contradiction. Thus, $\nu[a, r) = 0$.

Finally, set $N := (-\infty, q)$ in case (a), $N := (a, q)$ in case (b), and $N := [a, r)$ in case (c). Define $\nu \in \bar{\mathcal{C}}$ by density (1.55). Then, $\mu$-almost surely $\mathbf{1}_N + \frac{d\nu}{d\mu} > 0$. □

***Remark 1.4.17.*** *The empty set is a maximal common nullset, if in Proposition 1.4.16 condition (1) holds and $\mu(-y, q) = 0$ or if condition (4)(b) holds and $\mu(a, q) = 0$.*

***Remark 1.4.18.*** *Proposition 1.4.16 covers all cases which are considered in Proposition 1.4.15.*

*Proof.* 1.4.16(1)&(2) cover the case 1.4.15(3)(a). 1.4.16(2) covers the case 1.4.15(3)(b). 1.4.16(3) covers the case 1.4.15(3)(c).

Finally, suppose that 1.4.15(3)(d) holds, but 1.4.15(3)(c) is not satisfied. If $\lambda \neq d$ and $r > a$, then

$$-y \;>\; \frac{\lambda - d}{\lambda} \cdot a \;+\; \frac{d}{\lambda} \cdot q.$$

If $\mu(r, q) > 0$, then $\bar{a} = \sup\{x \in \operatorname{supp} \mu : \; x < q\} > r$, thus 1.4.15(3)(c) holds, a contradiction. Hence, $\mu(r, q) = 0$. Altogether, it follows that 1.4.16(3)&(4) together cover the cases 1.4.15(3)(c)&(d). □

As a corollary of Proposition 1.4.16 and Theorem 1.4.13 we finally obtain a characterization of the solution.

***Corollary 1.4.19.*** *Assume that one of the equivalent conditions of Proposition 1.4.15 holds. Let $N$ be a maximal common nullset, cf. Propositions 1.4.16. $\nu$ is the unique minimizer of the relative entropy under the constraints (1.42), (1.43) and (1.44), if and only if its $\mu$-density is of the following form:*

$$\frac{d\nu}{d\mu}(x) = c \cdot \exp\left\{(h_1 + h_2 + h_3 x)\mathbf{1}_{(-\infty,q)}(x) + \left(h_2 + h_3 \frac{d}{d+u} x\right) \mathbf{1}_{\{q\}}(x)\right\} \cdot \mathbf{1}_{N^c}(x). \tag{1.56}$$

*Here, $c > 0$ is a normalizing constant, $h_1, h_2, h_3 \in \mathbb{R}$, and conditions (1.42), (1.43) and (1.44) need to be satisfied.*

*In particular, the minimizer $\nu$ is equivalent to $\mu$, if and only if the empty set is a maximal common nullset.*

*Proof.* The proof follows directly from Theorem 1.4.13. □

If $\nu$ is the minimizing density characterized in Corollary 1.4.19, the minimal relative entropy is given by the expression

$$\begin{aligned} H(\nu|\mu) &= \log c + \int (h_1 + h_2 + h_3 x) \mathbf{1}_{(-\infty,q)}(x)\, \nu(dx) \\ &\quad + \int \left( h_2 + h_3 \frac{d}{d+u} x \right) \mathbf{1}_{\{q\}}(x)\, \nu(dx) \\ &= \log c + h_1(\lambda - d) + h_2(\lambda + u) - h_3 \lambda y. \end{aligned}$$

**The Original Problem II: the General Case**

As we have already discussed before, the solution of the entropy minimization problem on the constraint set $\mathcal{C}$ can be obtained by minimizing over the solutions of the minimization problems under linear constraints.

***Proposition 1.4.20.*** *There exists a solution to the minimization problem of the relative entropy on $\mathcal{C}$, if and only if $a < -y < b$, or $-y$ is an atom of $\mu$. The collection of all solutions is given by (1.45).*

*Proof.* According to Proposition 1.4.3, a solution with finite relative entropy exists, if and only if $a < -y < b$, or $-y$ is an atom of $\mu$. Clearly, all solutions are obtained by solving first the minimization problem under constraints (1.42), (1.43) and (1.44) for fixed $q \in \mathbb{R}$, $d \in [0, \lambda]$ and $u \in [0, 1-\lambda]$ and then by minimizing over $q, u, d$. □

## 1.5 Entropy Minimization under a Shortfall Risk Constraint

In the current section, we consider a second example of the entropy minimization problem under a risk measure constraint: we discuss a shortfall risk constraint. As average value at risk, shortfall risk has many appealing properties. It is distribution-invariant, coherent and sensitive to the size of losses. In contrast to average value at risk, shortfall risk can be used for the consistent dynamic evaluation of financial positions, cf. Chapter 2.

**The minimization problem** Let $\ell : \mathbb{R} \to \mathbb{R}$ be a convex loss function, $z \in \mathbb{R}$ be a point in the interior of the range of $\ell$, and $\rho$ be shortfall risk associated with $\ell$ and $z$. Fix a reference probability $\mu \in \mathcal{M}_{1,c}$ with compact support and a constant $y \in \mathbb{R}$. We are interested in the problem of minimizing $H(\nu|\mu)$ where $\nu \in \mathcal{M}_{1,c}$ and $\rho'(\nu) = y$. We set $a := \inf\{x \in \mathbb{R} :\ x \in \text{supp}\ \mu\}$, and $b := \sup\{x \in \mathbb{R} :\ x \in \text{supp}\ \mu\}$. As in the case of average value at risk, we may restrict our attention to the constraint set

$$\mathcal{C} := \{\nu \in \mathcal{M}_1([a,b]) : \ \rho'(\nu) = y\}.$$

### 1.5.1 Existence of Solutions

A necessary and sufficient criterion for the existence of solutions can be formulated in terms of the parameters $a$, $b$ and $y$. The derivation is based on Lemma 1.4.1.

***Remark 1.5.1.*** *Since $\ell$ is convex, continuous and increasing, and $z$ is an element of the interior of the range of $\ell$, there exists a unique element $\ell^{-1}(z) \in \mathbb{R}$ such that $\ell(\ell^{-1}(z)) = z$. Moreover, $\ell$ is strictly increasing on $(\ell^{-1}(z) - \epsilon, \infty)$ for some $\epsilon > 0$.*

***Proposition 1.5.2.*** *Let $X \in L^\infty$. Then $\rho(X) = y$, if and only if $\int \ell(-X - y)dP = z$.*

*Proof.* Fix $X \in L^\infty$, and define $L : m \mapsto \int \ell(-m - X)dP$. Since $X \in L^\infty$ and $\ell$ is continuous, $L$ is continuous by the bounded convergence theorem. $L$ is decreasing, since $\ell$ is increasing. The fact that $z$ is an element of the interior of the range of $\ell$ implies that $\lim_{m\to-\infty} L(m) > z$, $\lim_{m\to\infty} L(m) < z$. Thus, there exists $y \in \mathbb{R}$ that solves $L(y) = z$. The solution is unique, since $\ell$ is strictly increasing on $(l^{-1}(z) - \epsilon, \infty)$ for some $\epsilon > 0$. Since $L(m) > z$ for all $m < y$, we obtain

$$\rho(X) = \inf\{m \in \mathbb{R} : \ L(m) \le z\} = y.$$

□

The following proposition characterizes the existence of solutions.

***Proposition 1.5.3.*** *The following conditions are equivalent.*

*(1) There exists $\nu \in \mathcal{C}$ such that $H(\nu|\mu) < \infty$.*

*(2) The minimal value of the relative entropy on $\mathcal{C}$ is finite and attained for some element of $\mathcal{C}$.*

*(3) $a$ is an atom of $\mu$ and $\ell(-a - y) = z$, or $b$ is an atom of $\mu$ and $\ell(-b - y) = z$, or*

$$\ell(-b - y) < z < \ell(-a - y). \tag{1.57}$$

*If (1.57) holds, then there exists $\nu \approx \mu$, $\nu \in \mathcal{C}$ with $H(\nu|\mu) < \infty$.*

*Proof.* (1) and (2) are clearly equivalent by Lemma 1.4.1. Assume now that (1) holds. Suppose that neither $a$ nor $b$ are atoms of $\mu$. Then neither $a$ nor $b$ are atoms of $\nu$. Since

$\nu \in \mathcal{C}$, we have supp $\nu \subseteq [a,b]$. If $\ell(-a-y) < z$, then $\ell(-x-y) < z$ for $x \in (a,b]$. If $\ell(-a-y) = z$, then $\ell(-x-y) < z$ for $x \in (a,b]$, since $\ell^{-1}\{z\}$ is a singleton and $\ell$ is increasing. Thus, if $\ell(-a-y) \le z$, then $\int \ell(-x-y)\nu(dx) < z$, since $a$ is not an atom of $\nu$. By Proposition 1.5.2 $y \neq \rho'(\nu)$, a contradiction. Thus, $\ell(-a-y) > z$. Analogously, one can show that $\ell(-b-y) < z$. If $a$ or $b$ are atoms, then additionally $\ell(-a-y) = z$ or $\ell(-b-y) = z$ is possible. This proves that (1) implies (3).

Conversely, assume that (3) holds. If $a$ is an atom of $\mu$ and $\ell(-a-y) = z$, define $\nu \ll \mu$ by $\frac{d\nu}{d\mu} = \frac{1}{\mu\{a\}}\mathbf{1}_{\{a\}}$. Then, $H(\nu|\mu) < \infty$ and $\int \ell(-x-y)\nu(dx) = \ell(-a-y) = z$, thus $\rho'(\nu) = y$ by Proposition 1.5.2. Analogously, if $b$ is an atom of $\mu$ and $\ell(-b-y) = z$, then $\frac{d\nu}{d\mu} = \frac{1}{\mu\{b\}}\mathbf{1}_{\{b\}}$ defines $\nu \in \mathcal{C}$ with $H(\nu|\mu) < \infty$.

Finally, assume that $\ell(-b-y) < z < \ell(-a-y)$. Since $\ell^{-1}\{z\}$ is a singleton, we obtain $\ell(-x-y) < z$ for $x > q$, $\ell(-x-y) > z$ for $x < q$, where $q := -\ell^{-1}(z) - y \in (a,b)$. Thus,

$$\begin{aligned} u_+ &:= \int \mathbf{1}_{[a,q]}(x)\ell(-x-y)\mu(dx) \ > \ z, \\ u_- &:= \int \mathbf{1}_{(q,b]}(x)\ell(-x-y)\mu(dx) \ < \ z. \end{aligned}$$

Choose $\alpha \in (0,1)$ such that

$$\alpha u_+ + (1-\alpha)u_- \ = \ z.$$

Then, we define $\nu \ll \mu$ by

$$\frac{d\nu}{d\mu} \ = \ \alpha \mathbf{1}_{[a,q]} + (1-\alpha)\mathbf{1}_{(q,b]}.$$

Clearly, $H(\nu|\mu) < \infty$ and $\nu \in \mathcal{C}$ by Proposition 1.5.2. $\square$

### 1.5.2 Structure of the Solution

Since shortfall risk imposes a linear constraint, the solution to the minimization problem follows directly from Theorem 1.4.5.

***Corollary 1.5.4.*** *Assume that (1.57) holds. Then $\nu$ is the unique minimizer of the relative entropy on $\mathcal{C}$, if and only if its $\mu$-density is of the following form:*

$$\frac{d\nu}{d\mu}(x) = c \cdot \exp\left(h \cdot \ell(-x-y)\right). \tag{1.58}$$

*Here, $c > 0$ is a normalizing constant, $h \in \mathbb{R}$, and the following conditions need to be satisfied:*

$$z = c \cdot \int \ell(-x-y) \exp\left(h \cdot \ell(-x-y)\right) \mu(dx), \tag{1.59}$$

$$1 = c \cdot \int \exp\left(h \cdot \ell(-x-y)\right) \mu(dx). \tag{1.60}$$

*Proof.* By Proposition 1.5.3 there exists $\nu \in \mathcal{C}$ with $H(\nu|\mu) < \infty$ and $\nu \approx \mu$. It follows from Theorem 1.4.5 that there exists a unique minimizer of the relative entropy. By Proposition 1.5.2, a measure $\nu \in \mathcal{M}_1([a,b])$ is an element of $\mathcal{C}$, if and only if $\int \ell(-x-y)\nu(dx) = z$. Thus, the minimizer $\nu$ has $\mu$-density (1.58) by Theorem 1.4.5. (1.59) is required by the constraint, (1.60) is a normalization. □

If $\nu$ is the minimizing density characterized in Corollary 1.5.4, then the minimal relative entropy is given by the expression

$$H(\nu|\mu) = \log c + h \cdot z.$$

**Remark 1.5.5.** *If $a$ is an atom of $\mu$ and $\ell(-a-y) = z$, or if $b$ is an atom of $\mu$ and $\ell(-b-y) = z$, then $\mathcal{C} = \{\delta_x\}$ with $x = a$ or $x = b$, respectively. Here, $\delta_x$ denotes the Dirac measure on $x \in \mathbb{R}$. Then the minimizer $\nu$ of the relative entropy is trivially unique, and $H(\nu|\mu) = -\log \mu\{x\}$ with $x = a$ or $x = b$, respectively.*

# Chapter 2

# Distribution-Invariant Dynamic Risk Measures

## 2.1 Introduction

The quantification of the risk of financial positions is a key task for both financial institutions and supervising authorities. Risk management and financial regulation relies on the proper assessment of downside risk. Since traditional approaches – such as *value at risk* – do in general not encourage diversification of positions, alternative risk measures need to be designed and investigated. In the context of static financial positions economically meaningful axioms were proposed in the seminal paper by Artzner et al. (1999). The original definition has been relaxed in many directions, and various robust representation results for risk measures have been obtained (see e.g. Föllmer and Schied (2002a), Föllmer and Schied (2002b), Delbaen (2002)). Risk measures for topological vector spaces were considered by Jaschke and Küchler (2001) and Frittelli and Rosazza (2002). For excellent overviews on static risk measures, we refer to Föllmer and Schied (2002c), Delbaen (2000) and Scandolo (2003).

While the theory of static risk measures is already well developed, sophisticated risk management and financial regulation require *dynamic* risk measures for *dynamic* financial positions. Monetary measures of downside risk must evaluate the total risk of both the terminal and all intermediate cash flows. The measurements must consistently be updated, as new information becomes available. In the current chapter, we suggest an axiomatically well-founded model for dynamic risk measures of dynamic cash flows in discrete time. As in the static case, the measurements can be interpreted as capital requirements that must be invested in a risk-free financial instrument until a terminal date.

For certain dynamic risk measures we prove a simple representation theorem in terms of static distribution-invariant risk measures. Besides standard conditions known from the static case, the essential axioms are roughly the following:

(1) Agents have access to a market of risk-free bonds. The risk of two positions is equal at the current date, if these can completely be transformed into each other by trading in the bond market in the future.

(2) Whether or not a terminal position has positive risk, depends only on its conditional distribution.

We propose two notions of dynamic consistency for such risk measures, namely acceptance and rejection consistency. We call a dynamic risk measure acceptance consistent (resp. rejection consistent), if it satisfies the following condition: If a position is acceptable (resp. not acceptable) in the future for sure, then it is acceptable (resp. not acceptable) today. It is shown that dynamic consistency is closely related to properties of the acceptance and rejection sets of the representing static risk measures. Here, we use the concept of measure convex sets known from Choquet theory. We completely characterize the class of static risk measures that corresponds to consistent dynamic risk measures.

Finally, we further investigate these static distribution-invariant risk measures. Both their acceptance and their rejection sets are convex subsets of the space of probability measures. This has a natural interpretation in the context of static financial positions. If two financial positions or lotteries are acceptable (resp. rejected), than any compound lottery that randomizes over the positions is again acceptable (resp. rejected). Under additional topological conditions, we prove that risk measures with such acceptance and rejection sets coincide exactly with the well-known shortfall risk, if they are convex in the sense of Föllmer and Schied (2002c). This result can then be applied to dynamically consistent, convex risk measures.

There are many ways to introduce risk measures in a dynamic setting. Most approaches in the literature generalize the static results on coherent or convex risk measures. In contrast, we focus on distribution invariance and the connection between dynamic consistency and measure convexity. This implies the close link between shortfall risk on the one hand, and dynamic consistency, convexity and distribution-invariance on the other hand.

The axiomatic approach of Riedel (2002) is related to the current chapter. He analyzes dynamic coherent risk measures for financial positions on a finite probability space. Under a strong dynamic consistency axiom, he obtains a robust representation of coherent, dynamically consistent risk measures. The notions of dynamic consistency in the context of risk measures go back to Wang (1996) and Wang (1999).

Artzner et al. (2003) consider financial processes as random variables on an extended state space including dates in time. This allows them to employ the standard approach for static coherent risk measures and to obtain a robust representation. They establish a connection between time consistency, stability of test probabilities and Bellman's principle, see also Delbaen (2003). The approaches of Riedel (2002) and Artzner et al. (2003) are related to the analysis of multiple priors in decision theory, see e.g. Epstein and Schneider (2003). Convex and coherent risk measures for continuous-time processes are investigated by Cheridito et al. (2003). An axiomatic analysis of convex, conditioned risk measures can be found in Detlefsen (2003) and Scandolo (2003).

In the current chapter, we impose a special type of distribution invariance on dynamic risk measures. In the static context, coherent and convex distribution-invariant risk measures have been investigated by Kusuoka (2001), Carlier and Dana (2003), and Kunze (2003). These can be represented in terms of robust mixtures of average value at risk or upper envelops of Choquet integrals with respect to distortions of probability measures.

The chapter is organized as follows. In Section 2 we propose an axiomatic characterization of dynamic risk measures. In Section 3, we investigate static risk measures considered as functionals on the space of probability measures, and prove a simple representation theorem for dynamic risk measures in terms of static risk measures. Dynamic consistency conditions and locally measure convex sets of probability measures are discussed in Section 4. In Section 5 we investigate the close link of dynamic consistency and shortfall risk. Section 6 concludes.

## 2.2 An Axiomatic Characterization of Risk Dynamics

We consider time periods $t = 0, 1, \dots, T$. The state space $(\Omega, \mathcal{F}, P)$ is a standard Borel probability space. $(\mathcal{F}_t)_{t=0,1,\dots,T}$ denotes a filtration, modelling the flow of information. We assume that at time 0 information is trivial, i.e. $\mathcal{F}_0 = \{\emptyset, \Omega\}$, and that at time $T$ all information is revealed, i.e. $\mathcal{F}_T = \mathcal{F}$.

We intend to construct an axiomatically well-founded model for dynamic risk of financial positions. A dynamic monetary measure of risk is a sequence of mappings $\rho = (\rho_t)_{t=0,1,\dots,T-1}$ evaluating the risk of dynamic cash flows or financial positions $D = (D_t)_{t=0,1,\dots,T}$. The quantity $\rho_t(D)$ is interpreted as a measure of the risk of position $D$ at time $t$. We suppose that the space of financial positions equals

$$\mathcal{D} = \{(D_t)_{t=0,1,\dots,T} : \ D_t \in L^\infty(\Omega, \mathcal{F}_t, P)\}.$$

The financial position that pays 1 at time $t$ for sure and 0 else will be denoted by

$$e_t = (0,\ 0,\ \dots,\ 0, \underbrace{1}_{t},\ 0, \dots,\ 0).$$

We assume that agents have access to a market of zero coupon bonds with maturity $T$. The price of a bond at time $t$ is given by an $\mathcal{F}_t$-measurable random variable $P_t^T$. Here $P_T^T \equiv 1$, that is, the bond is default free. Considering only a finite time horizon $T$, we suppose that bond prices are both bounded from below and above, i.e. $P_s^T \in [\epsilon, c]$ for some $0 < \epsilon < c < \infty$. We abstract from trading costs.

### 2.2.1 The Axioms

We will assume that a *dynamic risk measure* satisfies the following axioms for all $t = 0, 1, \dots, T-1$.

**A Adaptedness, Monotonicity and Invariance**

(1) *Adaptedness and Boundedness*:
$\rho_t(D) \in L^\infty(\Omega, \mathcal{F}_t, P)$

(2) *Inverse Monotonicity*:
If $D \geq D'$, then $\rho_t(D) \leq \rho_t(D')$.

(3) *Translation-invariance*:
If $Z \in L^\infty(\Omega, \mathcal{F}_t, P)$, then

$$\rho_t\left(D + \frac{Z}{P_t^T} \cdot e_T\right) = \rho_t(D) - Z.$$

A1 ensures that the risk $\rho_t(D)$ of a position $D$ evaluated at time $t$ depends only on information available at time $t$ (adaptedness). Since the position $D$ is bounded, it is reasonable that its risk is also bounded. A2 states that the downside risk of a position decreases, if the payoff of the position increases in all possible scenarios $\omega \in \Omega$.

The axiom of translation-invariance, A3, formalizes the idea that $\rho_t(D)$ is a capital requirement. If an investor invests an amount of $Z$ at time $t$ in a risk-free way until maturity $T$, her risk is reduced exactly by $Z$. In particular, A3 implies that

$$\rho_t\left(D + \frac{\rho_t(D)}{P_t^T} \cdot e_T\right) = 0.$$

We will interpret $\rho_t(D)$ as the monetary amount that should be added to $D$ at time $t$ and invested in risk-free bonds until the final date to make the position acceptable from the point of view of an investor or regulator, given the information at time $t$. A position $D$ is acceptable at time $t$, if its risk $\rho_t(D) \leq 0$. In this case, no positive monetary amount has to be added to the position to make it acceptable.

**B Independence of the past**

If $D_s = D'_s$ for all $s > t$, then $\rho_t(D) = \rho_t(D')$.

B captures the idea that *'sunk costs are sunk.'* When assessing the risk of a position $D \in \mathcal{D}$ at time $t$, only the future payoffs are taken into account.

**C Invariance under adapted transforms**

Let $t < u \leq T$, and assume that $Z \in L^\infty(\Omega, \mathcal{F}, P)$ is $\mathcal{F}_u$-measurable. Then

$$\rho_t(D + Z \cdot P_u^T \cdot e_u - Z \cdot e_T) = \rho_t(D).$$

Axiom C can be interpreted as follows. An agent holding a financial position $D$ can form a contingent plan to transform $D$ into $D' = D + Z \cdot P_u^T \cdot e_u - Z \cdot e_T$ without facing any risk at time $u$:

- Sell $Z$ zero-coupon bonds at time $u$.
- Pay $Z$ to the bond owners at time $T$.

Vice versa, an agent holding $D'$ can form a contingent plan to transform $D'$ into $D$ without facing any risk at time $u$ by following the reversed strategy. For the agent the realization of these contingent plans is clearly feasible at the current date $t$, but it is also still feasible at the later date $t+1$, since $u$ is *strictly* bigger than $t$. Hence, both positions $D$ and $D'$ are equivalent for the agent at least until date $t+1$. Thus, *before* time $t+1$ they should have the same risk. In particular, the relation $\rho_t(D) = \rho_t(D')$ should hold.

From the viewpoint of a regulator the same reasoning applies. It is not necessary to impose different monetary requirements on the positions $D$ and $D'$ already at time $t$, if they can be transformed into each other at a later date without incurring any cost.

***Remark 2.2.1.*** *In Axiom C we state that risk is invariant for positions that can be transformed into each other using zero-coupon bonds. One could argue that risk should also be invariant under a more general class of transformations involving possibly other* *financial instruments. Observe that such an approach would* add *more restrictions on the risk measure, thus* decrease *the level of generality of the analysis.*

***Definition 2.2.2.*** *A mapping* $\rho = (\rho_t)_{t=0,1,\dots,T-1} : \mathcal{D} \times \Omega \to \mathbb{R}^T$ *is a* dynamic risk measure *if it satisfies the axioms A1, A2, A3, B and C.*

### 2.2.2 Distribution-Invariance

Let $\rho$ be a dynamic risk measure.
We define the *acceptance indicator* $a = (a_t)_{t=0,1,\dots,T-1}$ of $\rho$ by

$$a_t(D)(\omega) := \mathbf{1}_{(-\infty,0]}(\rho_t(D)(\omega)).$$

If $a_t(D) = 1$, at date $t$ the risk of $D$ is less or equal to 0 and no positive monetary amount has to be added to $D$ to make it acceptable. Conversely, if $a_t(D) = 0$, a positive monetary amount must be added to $D$ to make the position acceptable at date $t$.

We denote by $\mathcal{M}_{1,c}(\mathbb{R})$ the space of *probability measures on the real line* with compact support. If $Y$ is a real-valued random variable defined on $(\Omega, \mathcal{F}, P)$, we denote by $\mathcal{L}(Y|\mathcal{F}_t)$ the *regular conditional distribution* of $Y$ given $\mathcal{F}_t$. We briefly recall the definition of regular conditional distributions and results regarding existence and uniqueness.

***Definition 2.2.3.*** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, and let* $Y$ *be a measurable function on* $\Omega$ *into any measurable space* $(T, \mathcal{B})$*. Let* $\mathcal{G}$ *be a sub-*$\sigma$*-algebra of* $\mathcal{F}$*. Then a regular conditional distribution* $\mathcal{L}(Y|\mathcal{G})$ *of* $Y$ *given* $\mathcal{G}$ *is defined as a function from* $\Omega \times \mathcal{B}$ *into* $[0,1]$ *such that*

*(1) for* $P$*-almost all* $\omega \in \Omega$, $\mathcal{L}(Y|\mathcal{G})(\omega, \cdot)$ *is a probability measure on* $\mathcal{B}$*.*

*(2) for each* $B \in \mathcal{B}$, $\mathcal{L}(Y|\mathcal{G})(\cdot, B)$ *is* $\mathcal{G}$*-measurable.*

*(3) for* $B \in \mathcal{B}$ *and for all* $C \in \mathcal{G}$ *it holds that*

$$\int_C \mathcal{L}(Y|\mathcal{G})(\omega, B) P(d\omega) = \int_C \mathbf{1}_{Y \in B}(\omega) P(d\omega).$$

***Theorem 2.2.4.*** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, and let* $Y$ *be a measurable function on* $\Omega$ *into any standard Borel space* $(T, \mathcal{B})$*. Let* $\mathcal{G}$ *be a sub-*$\sigma$*-algebra of* $\mathcal{F}$*. Then a regular conditional distribution* $\mathcal{L}(Y|\mathcal{G})$ *of* $Y$ *given* $\mathcal{G}$ *exists. It is unique in the following sense: If* $\hat{\mathcal{L}}(Y|\mathcal{G})$ *is another regular conditional distribution, then the two laws* $\mathcal{L}(Y|\mathcal{G})(\omega, \cdot)$ *and* $\hat{\mathcal{L}}(Y|\mathcal{G})(\omega, \cdot)$ *are equal for* $P$*-almost all* $\omega \in \Omega$*.*

We will now introduce a notion of distribution-invariance for risk measures.

***Definition 2.2.5.*** *The dynamic risk measure $\rho$ is called* distribution-invariant at maturity *or* M-invariant *if there exists a measurable mapping*

$$H_t : \mathcal{M}_{1,c}(\mathbb{R}) \to \{0,1\}$$

*such that for all terminal positions $D = D_T \cdot e_T \in \mathcal{D}$*

$$a_t(D) = H_t(\mathcal{L}(D_T|\mathcal{F}_t)).$$

M-invariance formalizes the following idea. The purpose of a risk measure is to quantify the downside risk of a financial position. If a financial institution evaluates the risk of a fixed financial cash flow $Z$ to be paid at a fixed reference point in time $T$, and if the distribution is known, it is reasonable to assume that acceptability should depend only on the conditional distribution of $Z$ given the present information. The use of conditional distributions formalizes the idea that information is processed in a Bayesian fashion.

Of course, if we do *not* fix $Z$ assuming instead that $Z$ is invested into some financial asset or that $Z$ is a position in a larger portfolio, *then* total risk is determined by the conditional distributions and the dependence structure of *all* financial random variables involved. But, if we would like to evaluate a *fixed $Z$ alone*, downside risk should be understood as a property of its conditional distribution only.

## 2.3 Representation of Distribution-Invariant Risk

Dynamic M-invariant risk measures can be represented in terms of static distribution-invariant risk measures. This fact is indeed not surprising, and we will state the exact result in Theorem 2.3.9. The result is useful for the construction of examples of dynamic risk measures. Moreover, dynamic consistency which will be investigated in Section 2.4 can be characterized via properties of the representing static risk measures.

### 2.3.1 Static Distribution-Invariant Risk Measures

Most of the literature on static and dynamic risk measures focuses on coherence and convexity. In such a context it is useful to define risk measures as functionals on a space of financial positions. In contrast, in the current chapter issues like distribution-invariance and dynamic consistency are crucial, and it will be convenient to interpret static distribution-invariant risk measures as functionals on probability measures. On the space $\mathcal{M}_{1,c}(\mathbb{R})$ of probability measures on the real line with compact support a partial order $\leq$ is given by *stochastic dominance.*

***Definition 2.3.1.*** *A mapping $\Theta : \mathcal{M}_{1,c}(\mathbb{R}) \to \mathbb{R}$ is called a* risk measure *if it satisfies the following conditions for all $\mu, \nu \in \mathcal{M}_{1,c}(\mathbb{R})$:*

- Inverse Monotonicity: *If $\mu \leq \nu$, then $\Theta(\mu) \geq \Theta(\nu)$.*
- Translation-Invariance: *If $m \in \mathbb{R}$, then $\Theta(T_m\, \mu) = \Theta(\mu) - m$.*
  *Here, for $m \in \mathbb{R}$ the translation operator $T_m$ is given by $(T_m\mu)(\cdot) = \mu(\cdot - m)$.*

Inverse monotonicity captures the intuition that risk decreases, if a financial position is concentrated on larger values. Translation invariance formalizes the idea that the risk of a position is actually a monetary requirement: if a monetary amount $m$ is added to the position $\mu$, its risk is reduced by the same amount.

We introduced static risk measures as functionals on the space of probability measures on the real line, while the classical literature on risk measures investigates functionals on spaces of financial positions. The two notions are equivalent in the following sense:

Suppose that $(\Omega', \mathcal{F}', P')$ is an atomless probability space, and let $L^\infty(\Omega', \mathcal{F}', P')$ be a space of financial positions. If $\Theta : \mathcal{M}_{1,c}(\mathbb{R}) \to \mathbb{R}$ is a risk measure in the sense of Definition 2.3.1, then $\Theta'(X) = \Theta(\mathcal{L}(X))$ defines a distribution-invariant risk measure on $L^\infty(\Omega', \mathcal{F}', P')$. Conversely, if $\Theta'$ is a distribution-invariant risk measure on $L^\infty(\Omega', \mathcal{F}', P')$, then $\Theta(\mu) = \Theta'(X)$ for some $X \sim \mu$ defines a risk measure on $\mathcal{M}_{1,c}(\mathbb{R})$ in the sense of Definition 2.3.1.

This identification helps to derive properties of risk measures on $\mathcal{M}_{1,c}(\mathbb{R})$ from the classical case. We will now show that any risk measure on $\mathcal{M}_{1,c}(\mathbb{R})$ is Lipschitz-continuous with respect to a particular Vasserstein metric. This implies, in particular, that risk measures are measurable functionals with respect to the Borel-$\sigma$-algebra of the weak topology.

***Lemma 2.3.2.*** *Any risk measure $\Theta : \mathcal{M}_{1,c}(\mathbb{R}) \to \mathbb{R}$ is Lipschitz continuous with respect to the Vasserstein distance $V_\infty$:*

$$|\Theta(\mu) - \Theta(\nu)| \leq V_\infty(\mu, \nu).$$

*Here, for $\mu, \nu \in \mathcal{M}_{1,c}(\mathbb{R})$ the Vasserstein distance is defined by*

$$V_\infty(\mu, \nu) = \inf \|X - Y\|,$$

*where $\|\cdot\|$ denotes the essential supremum and the infimum is taken over all pairs of random variables $X \sim \mu$ and $Y \sim \nu$ on some atomless probability space.*

*Proof.*
Let $\mu, \nu \in \mathcal{M}_{1,c}(\mathbb{R})$ be given. Assume that $X \sim \mu$, $Y \sim \nu$ and $X, Y \in L^\infty(\Omega', \mathcal{F}', P')$ for some probability space $(\Omega', \mathcal{F}', P')$. W.l.o.g we may assume that $(\Omega', \mathcal{F}', P')$ is atomless by identifying every atom with a subinterval of $((0,1), \lambda)$ of appropriate length; this does neither change the joint distribution of $(X, Y)$ nor the norm $\| X - Y \|$. Then by the Lipschitz continuity of $\Theta'$ it follows that

$$|\Theta(\mu) - \Theta(\nu)| = |\Theta'(X) - \Theta'(Y)| \leq \|X - Y\|.$$

Note that the Lipschitz continuity of $\Theta'$ is a trivial consequence of the monotonicity and translation invariance of $\Theta'$, cf. Lemma 4.3 in Föllmer and Schied (2002c).
This implies the claim. □

***Remark 2.3.3.*** *For measures on $\mathbb{R}$ the Vasserstein metric $V_\infty$ can be represented in terms of the inverse of the distribution functions (i.e. the quantile functions) of the measures $\mu, \nu \in \mathcal{M}_{1,c}(\mathbb{R})$, cf. Owen (1987). We denote by $F_\mu^{-1}$ and $F_\nu^{-1}$ the right-continuous inverse of the distribution function of $\mu$ and $\nu$, respectively. It holds that*

$$V_\infty(\mu, \nu) = \sup_{0<u<1} |F_\mu^{-1}(u) - F_\nu^{-1}(u)|. \tag{2.1}$$

*For other Vasserstein metrics see Owen (1987) and Rachev (1991).*

**Lemma 2.3.4.** *The $V_\infty$-metric generates the Borel-$\sigma$-algebra on $\mathcal{M}_{1,c}(\mathbb{R})$ induced by the weak topology.*

*Proof.* The quantile function

$$F_\mu^{-1}(u) = q_\mu(u) = \sup\{x : \mu(-\infty, x) \leq u\} \tag{2.2}$$

is product measurable on $\mathcal{M}_{1,c}(\mathbb{R}) \times [0,1]$, since the set $\{(\mu, u) : \mu(-\infty, x) \leq u\}$ is measurable for each $x$ and the supremum in (2.2) can be restricted to rational $x$. More precisely, the product measurability is implied by the following identities: for any $z \in \mathbb{R}$ it holds that

$$\{(\mu, u) : q_\mu(u) \geq z\} = \left\{(\mu, u) : \sup_{x \in \mathbb{Q}}\{x : \mu(-\infty, x) \leq u\} \geq z\right\}$$
$$= \bigcap_{x \in \mathbb{Q},\, x<z} \{(\mu, u) : \mu(-\infty, x) \leq u\}.$$

Now fix $\mu \in \mathcal{M}_{1,c}(\mathbb{R})$. Then the Vasserstein ball $\{\nu : V_\infty(\mu, \nu) < \epsilon\}$ is measurable with respect to the standard $\sigma$-algebra, since the supremum in (2.1) can be restricted to rational $u$. More precisely, for $u \in [0,1]$ the function $q_\cdot(u) : \mathcal{M}_{1,c}(\mathbb{R}) \to \mathbb{R}$ is measurable with respect to the standard $\sigma$-algebra. Hence,

$$V_\infty(\mu, \cdot) \;=\; \sup_{0<u<1,\, u\in\mathbb{Q}} |q_\mu(u) - q_\cdot(u)|$$

is measurable with respect to the standard $\sigma$-algebra. This implies the measurability of the Vasserstein ball.

Hence, the Borel-$\sigma$-algebra generated by the $V_\infty$-topology is coarser than the standard $\sigma$-algebra. The converse is true, since the Vasserstein topology is finer than the weak topology. □

***Corollary 2.3.5.*** *A risk measure $\Theta : \mathcal{M}_{1,c}(\mathbb{R}) \to \mathbb{R}$ is measurable with respect to the Borel-$\sigma$-algebra on $\mathcal{M}_{1,c}(\mathbb{R})$ generated by the weak topology.*

Acceptance sets on the level of probability distributions can be defined by

$$\mathcal{N}_\Theta = \{\mu \in \mathcal{M}_{1,c}(\mathbb{R}) : \; \Theta(\mu) \leq 0\}.$$

For any given risk measure, the acceptance set consists of all probability distributions with non positive risk. Conversely, as in the case of financial positions, acceptance sets may be used to define corresponding risk measures. The following lemma is a simple corollary of the well-known results on classical risk measures, see e.g. Propositions 4.5 and 4.6 in Föllmer and Schied (2002c).

***Lemma 2.3.6.*** *Assume that $\mathcal{N} \subseteq \mathcal{M}_{1,c}(\mathbb{R})$ is non-empty, and satisfies the following two conditions:*

$$\inf\{m \in \mathbb{R} : \quad \delta_m \in \mathcal{N}\} > -\infty. \tag{2.3}$$

$$\mu \in \mathcal{N},\; \nu \in \mathcal{M}_{1,c}(\mathbb{R}),\; \nu \geq \mu \;\Rightarrow\; \nu \in \mathcal{N}. \tag{2.4}$$

*Then $\mathcal{N}$ induces a risk measure $\Theta$ by*

$$\Theta(\mu) = \inf\{m \in \mathbb{R} : \quad T_m(\mu) \in \mathcal{N}\}.$$

*$\mathcal{N}$ is included in the acceptance set of $\Theta$.*

Recall that the measure of risk $\Theta'$ on the space $L^\infty(\Omega', \mathcal{F}', P')$ is called *convex*, if $\Theta'(\alpha X + (1-\alpha)Y) \leq \alpha\Theta'(X) + (1-\alpha)\Theta'(Y)$ for all $X, Y \in L^\infty(\Omega', \mathcal{F}', P')$, $\alpha \in [0,1]$. $\Theta'$ is called *positively homogenous*, if $\Theta'(\lambda X) = \lambda\Theta'(X)$ for all $X \in L^\infty(\Omega', \mathcal{F}', P')$ and $\lambda \geq 0$. The risk measure is *coherent*, if it is both convex and positively homogenous. In the next definition we introduce the notions of convexity and coherence for risk measures on $\mathcal{M}_{1,c}(\mathbb{R})$ employing the correspondence to the classical case.

***Definition 2.3.7.*** *Let $\Theta$ and $\Theta'$ be risk measures as defined above. We say that $\Theta$ is* convex *(resp.* coherent*) if $\Theta'$ is convex (resp. coherent).*

***Lemma 2.3.8.*** *The notions of* convexity *and* coherence *of risk measures on $\mathcal{M}_{1,c}(\mathbb{R})$ are well-defined.*

*Proof.* We have to show that both concepts do not depend on the choice of the atomless probability space $(\Omega', \mathcal{F}', P')$. Let $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ be another atomless probability space, and let $Z$ be a unif$(0,1)$-distributed random variable on $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$. Let $X', Y' : \Omega' \to \mathbb{R}$ be two random variables. By Borel's theorem (see e.g. Theorem 2.19 in Kallenberg (1997)) it follows that there exists a measurable mapping $(g_1, g_2) : [0,1] \to \mathbb{R}^2$ such that $(g_1 \circ Z, g_2 \circ Z) \sim (X', Y')$. We set $\hat{X} = g_1 \circ Z$, $\hat{Y} = g_2 \circ Z$.

Now suppose that $\alpha \in (0,1)$, and for random variables $\hat{X}, \hat{Y} : \hat{\Omega} \to \mathbb{R}$,

$$\Theta(\mathcal{L}(\alpha\hat{X} + (1-\alpha)\hat{Y}) \leq \alpha\Theta(\mathcal{L}(\hat{X})) + (1-\alpha)\Theta(\mathcal{L}(\hat{Y})).$$

Let $\alpha \in (0,1)$, and random variables $X', Y' : \Omega' \to \mathbb{R}$ be given. Then there exists random variables $\hat{X}, \hat{Y} : \hat{\Omega} \to \mathbb{R}$ such that $(X', Y') \sim (\hat{X}, \hat{Y})$. We obtain

$$\begin{aligned} \Theta(\mathcal{L}(\alpha X' + (1-\alpha)Y') &= \Theta(\mathcal{L}(\alpha\hat{X} + (1-\alpha)\hat{Y}) \\ \leq \alpha\Theta(\mathcal{L}(\hat{X})) + (1-\alpha)\Theta(\mathcal{L}(\hat{Y})) &= \alpha\Theta(\mathcal{L}(X')) + (1-\alpha)\Theta(\mathcal{L}(Y')). \end{aligned}$$

The same implication holds if we reverse the roles of $\Omega'$ and $\hat{\Omega}$. It follows that the definition of convexity of $\Theta$ does not rely on the choice of the probability space $(\Omega', \mathcal{F}', P')$. An analogous argument holds for coherence. □

Under additional continuity conditions, static distribution-invariant risk measures can be represented as robust mixtures of average value at risk and as upper envelops of Choquet integrals with respect to distortions of probability measures. Such characterizations of convex and coherent risk measures follow from results of Kusuoka (2001), Carlier and Dana (2003), and Kunze (2003).

### 2.3.2 A Simple Representation Theorem

The following representation characterizes M-invariant dynamic risk measures in a simple way. To keep the notation simple, we denote by

$$\mathcal{T}_t(D) := \mathcal{L}\left(\sum_{u=t+1}^{T} \frac{D_u}{P_u^T} \middle| \mathcal{F}_t\right)$$

the conditional distribution of a specific terminal position associated with $D \in \mathcal{D}$.

***Theorem 2.3.9.*** *Assume that the probability space is rich in the sense that there exists a unif(0,1)-distributed random variable independent of $\mathcal{F}_{T-1}$. Then an M-invariant dynamic risk measure $\rho$ can be represented by*

$$\rho_t(D) = P_t^T \cdot \Theta_t \left[\mathcal{T}_t(D)\right]. \tag{2.5}$$

*Here, $\Theta_t$ is a static risk measure considered as a functional on probability measures on $\mathbb{R}$, see Definition 2.3.1. The risk measures $\Theta_t$ in the representation are unique, and the acceptance set of $\Theta_t$ is given by*

$$\mathcal{N}_t = \{\mu \in \mathcal{M}_{1,c}(\mathbb{R}) : \; H_t(\mu) = 1\}. \tag{2.6}$$

*$H_t : \mathcal{M}_{1,c} \to \{0,1\}$ is the mapping introduced in Definition 2.2.5.*

*Proof.* Let $D \in \mathcal{D}$ be given. By independence of the past and invariance under adapted transforms we obtain

$$\rho_t(D) = \rho_t\left(\sum_{u=t+1}^{T} D_u \cdot e_u\right) = \rho_t\left(\sum_{u=t+1}^{T} \frac{D_u}{P_u^T} \cdot e_T\right)$$

Thus, w.l.o.g. we may assume that $D = K \cdot e_T$ with $K \in L^\infty(\Omega, \mathcal{F}, P)$.
For $t = 0, 1, \ldots, T-1$ we define the sets

$$\mathcal{N}_t = \{\mu \in \mathcal{M}_{1,c}(\mathbb{R}) : \; H_t(\mu) = 1\}.$$

We show that $\mathcal{N}_t$ induces a static risk measure.

First, we prove property (2.3): Let $M' \in L^\infty(\Omega, \mathcal{F}, P)$ be arbitrary. Define

$$M := M' + \frac{\rho_t(M' \cdot e_T) - 1}{P_t^T}.$$

By assumption, $P_t^T$ is bounded away from zero and $\rho_t(M' \cdot e_T) \in L^\infty(\Omega, \mathcal{F}, P)$. Thus, $M \in L^\infty(\Omega, \mathcal{F}, P)$. By translation invariance,

$$\rho_t(M \cdot e_T) = \rho_t\left(M' \cdot e_T + \frac{\rho_t(M' \cdot e_T) - 1}{P_t^T} \cdot e_T\right) = \rho_t(M' \cdot e_T) - \rho_t(M' \cdot e_T) + 1 > 0.$$

Let $m \in \mathbb{R}$, $m \leq -\|M\|_\infty$. By inverse monotonicity, $\rho_t(m \cdot e_T) \geq \rho_t(M \cdot e_T) > 0$. Hence,

$$H_t(\delta_m) = a_t(m \cdot e_T) = 0.$$

This implies that $\inf\{m \in \mathbb{R} : \; \delta_m \in \mathcal{N}_t\} > -\infty$.

Second, we prove property (2.4): Let $\mu \in \mathcal{N}_t$, $\nu \in \mathcal{M}_{1,c}(\mathbb{R})$, and $\nu \geq \mu$. Since the filtered probability space is rich, there exists a random variable $Z$ uniformly distributed on $(0,1)$ and independent of $\mathcal{F}_{T-1}$. Define $M := q_\mu(Z) \sim \mu$ and $N := q_\nu(Z) \sim \nu$, where $q_\mu$ and $q_\nu$ are the quantile functions of $\mu$ and $\nu$, respectively. Since $\nu$ stochastically dominates $\mu$, we have $N \geq M$. By monotonicity, $\rho_t(N \cdot e_T) \leq \rho_t(M \cdot e_T)$. This implies $H_t(\nu) = 1$, since $H_t(\mu) = 1$ by assumption. Hence, $\nu \in \mathcal{N}_t$.

We denote the static risk measure induced by the set $\mathcal{N}_t$ by $\Theta_t$ and have to show that

$$\rho_t(D) = P_t^T \cdot \Theta_t(\mathcal{L}(K|\mathcal{F}_t)).$$

By $T : \mathbb{R} \times \mathcal{M}_{1,c}(\mathbb{R}) \to \mathcal{M}_{1,c}(\mathbb{R})$ we denote the translation operator, i.e. $T_r\mu(A) = \mu(A - r)$ for $r \in \mathbb{R}$, $\mu \in \mathcal{M}_{1,c}(\mathbb{R})$ and measurable $A \subseteq \mathbb{R}$.
Since $\rho_t(D) \cdot \left(P_t^T\right)^{-1}$ is $\mathcal{F}_t$-measurable and bounded, we get

$$\frac{\rho_t(D)}{P_t^T} = \text{ess inf } \left\{ m \in L^\infty(\Omega, \mathcal{F}_t, P) : \ \frac{\rho_t(D)}{P_t^T} \leq m \right\}$$

Now let $m \in L^\infty(\Omega, \mathcal{F}_t, P)$ be arbitrary. By translation-invariance,

$$\frac{\rho_t(D)}{P_t^T} - m = \frac{\rho_t(D + m \cdot e_T)}{P_t^T}.$$

Thus,

$$\frac{\rho_t(D)}{P_t^T} \leq m \ \Leftrightarrow \ \rho_t(D + m \cdot e_T) \leq 0 \ \Leftrightarrow \ \mathcal{L}(K + m|\mathcal{F}_t) \in \mathcal{N}_t \ \Leftrightarrow \ T_m\mathcal{L}(K|\mathcal{F}_t) \in \mathcal{N}_t.$$

This implies

$$\frac{\rho_t(D)}{P_t^T} = \text{ess inf } \left\{ m \in L^\infty(\Omega, \mathcal{F}_t, P) : \ T_{m(\omega)}\mathcal{L}(K|\mathcal{F}_t)(\omega) \in \mathcal{N}_t \text{ for all } \omega \in \Omega \right\}$$

We have to show that the right hand side equals $\Theta_t(\mathcal{L}(K|\mathcal{F}_t))$:
First, observe that $\Theta_t : \mathcal{M}_{1,c}(\mathbb{R}) \to \mathbb{R}$ is Lipschitz continuous with respect to the Vasserstein metric $V_\infty$. This implies that $\hat{m} := \Theta_t(\mathcal{L}(K|\mathcal{F}_t)) \in L^\infty(\Omega, \mathcal{F}_t, P)$. Clearly, $T_{\hat{m}(\omega)}\mathcal{L}(K|\mathcal{F}_t)(\omega) \in \mathcal{N}_t$ for all $\omega \in \Omega$. Thus, $\hat{m} \geq \frac{\rho_t(D)}{P_t^T}$.
Second, let $m \in L^\infty(\Omega, \mathcal{F}_t, P)$ such that $T_{m(\omega)}\mathcal{L}(K|\mathcal{F}_t)(\omega) \in \mathcal{N}_t$ for all $\omega \in \Omega$. Since

$$\hat{m}(\omega) = \Theta_t(\mathcal{L}(K|\mathcal{F}_t)(\omega)) = \inf\{r \in \mathbb{R} : \ T_r\mathcal{L}(K|\mathcal{F}_t)(\omega) \in \mathcal{N}_t\},$$

we obtain in particular $\hat{m}(\omega) \leq m(\omega)$ for all $\omega \in \Omega$. Hence $\hat{m} \leq \frac{\rho_t(D)}{P_t^T}$.

Finally, we show that $\mathcal{N}_t$ is indeed the acceptance set of $\Theta_t$ and the uniqueness of the representation. Since the probability space is rich, for $\mu$ we can find $M \in L^\infty$ with $\mathcal{L}(M|\mathcal{F}_t) = \mu$. Uniqueness is implied by the equality

$$\Theta_t(\mu) = \frac{\rho_t(M \cdot e_T)}{P_t^T}.$$

Moreover, if $\Theta_t(\mu) \leq 0$, then $H_t(\mu) = 1$, thus $\mu \in \mathcal{N}_t$. This implies that $\mathcal{N}_t$ is indeed the acceptance set of $\Theta_t$. $\square$

***Remark 2.3.10.*** *In Theorem 2.3.9, Corollary 2.4.2, Theorem 2.4.4, and Theorem 2.4.5 we assume that the underlying probability spaces are rich in an appropriate sense. We formulate these requirements in terms of unif$(0,1)$-distributed random variables. This special assumption on the distribution is not necessary and can be relaxed. Instead, it is equivalent to assume the existence of an arbitrary* continuous *distribution.*

If the dynamic risk measure $\rho$ is *positively homogeneous*, i.e. $\rho_t(\alpha \cdot D) = \alpha \cdot \rho_t(D)$ for $\alpha \in L^\infty(\Omega, \mathcal{F}_t, P)$, $\alpha \geq 0$, then the representing measures $\Theta_t$ are *positively homogeneous* and the representation becomes:

$$\rho_t(D) = \Theta_t \left[ \mathcal{L} \left( \sum_{u=t+1}^{T} \frac{P_t^T}{P_u^T} \cdot D_u \middle| \mathcal{F}_t \right) \right].$$

If interest rates are deterministic, this representation of *positively homogeneous* risk measures involves only discounted positions. This parallels the results of Riedel (2002) on *coherent* dynamic risk measures on finite probability spaces.

The next lemma states the converse of Theorem 2.3.9: if the components of $\rho$ are defined as in (2.5), then $\rho$ is an M-invariant dynamic risk measure.

***Lemma 2.3.11.*** *Let $(\Theta_t)_{t=0,1,\dots,T-1}$ be a sequence of static risk measures as introduced in Definition 2.3.1. Then (2.5) defines an M-invariant dynamic risk measure.*

*Proof.* Adaptedness, inverse monotonicity, and independence of the past are immediate. Boundedness follows from the boundedness assumptions on the bond prices and the Lipschitz continuity of static risk measures with respect to the Vasserstein metric $V_\infty$.

We denote again by $T : \mathbb{R} \times \mathcal{M}_{1,c}(\mathbb{R}) \to \mathcal{M}_{1,c}(\mathbb{R})$ the translation operator. Then translation invariance can be verified as follows. Let $Z \in L^\infty(\Omega, \mathcal{F}_t, P)$. Then

$$\begin{aligned}
\rho_t \left( D + \frac{Z}{P_t^T} \cdot e_T \right) &= P_t^T \cdot \Theta_t \left( \mathcal{L} \left( \sum_{u=t+1}^{T} \frac{D_u}{P_u^T} + \frac{Z}{P_t^T} \middle| \mathcal{F}_t \right) \right) \\
&= P_t^T \cdot \Theta_t \left( T_{\frac{Z}{P_t^T}} \, \mathcal{L} \left( \sum_{u=t+1}^{T} \frac{D_u}{P_u^T} \middle| \mathcal{F}_t \right) \right) = P_t^T \cdot \Theta_t \left( \mathcal{L} \left( \sum_{u=t+1}^{T} \frac{D_u}{P_u^T} \middle| \mathcal{F}_t \right) \right) - Z \\
&= \rho_t(D) - Z
\end{aligned}$$

In order to prove invariance under adapted transforms let $t < v \leq T$, and assume that $Z \in L^\infty(\Omega, \mathcal{F}_v, P)$. Let $D \in \mathcal{D}$ be given, and define $D' = D + Z \cdot P_v^T \cdot e_v - Z \cdot e_T$. The claim follows by observing

$$\sum_{u=t+1}^{T} \frac{D_u}{P_u^T} = \sum_{u=t+1}^{T} \frac{D_u}{P_u^T} + \frac{Z \cdot P_v^T}{P_v^T} - Z = \sum_{u=t+1}^{T} \frac{D'_u}{P_u^T}$$

□

***Remark 2.3.12.*** *At a given time $t$ the positions $D \in \mathcal{D}$ and $\sum_{u=t+1}^{T} \frac{D_u}{P_u^T} \cdot e_T$ have the same risk. This is implied by axioms B and C, namely by invariance under adapted transforms and independence of the past. The risk of $D$ is then calculated by discounting the static risk of the conditional distribution of the terminal payment $\sum_{u=t+1}^{T} \frac{D_u}{P_u^T}$.*

*This result can be generalized in the following way. Instead of requiring axioms B and C, we could assume that at a given time $t$ the position $D \in \mathcal{D}$ has the same risk as a terminal position $K_t(D) \cdot e_T$, where $K_t(D) \in \mathcal{L}^\infty(\Omega, \mathcal{F}, P)$. Here, we suppose that on terminal positions the mapping $D \mapsto K_t(D) \cdot e_T$ is the identity. Define $\mathcal{T}_t(D) := \mathcal{L}(K_t(D)|\mathcal{F}_t)$. Then Theorem 2.3.9 is still true. If additionally the mappings $K_t$ are monotone increasing on $\mathcal{D}$, the same applies to Lemma 2.3.11 and the results of Sections 4 and 5. This generalization is important, if due to liquidity risk it is more expensive to transfer large negative amounts to the terminal date than small negative amounts.*

## 2.4 Dynamic Consistency

The axioms A, B, and C describe the properties of the components $\rho_t$ of the risk measure $\rho$, but do not require any consistency of risk evaluated at different dates. This fact is also apparent from Theorem 2.3.9 and Lemma 2.3.11: the representing static risk measures $\Theta_t$ can arbitrarily be chosen for different values of $t$. In this section we will investigate the implications of consistency requirements in time.

### 2.4.1 Representation of Consistent Risk Measures

***Definition 2.4.1.*** *A dynamic risk measure $\rho$ is*

- acceptance consistent*, if*

$$a_{t+1}(D) \equiv 1 \quad \Rightarrow \quad a_t(D - D_{t+1} \cdot e_{t+1}) \equiv 1,$$

- rejection consistent*, if*

$$a_{t+1}(D) \equiv 0 \quad \Rightarrow \quad a_t(D - D_{t+1} \cdot e_{t+1}) \equiv 0.$$

*Here, equality is always understood $P$-almost surely.*

*Acceptance consistency* captures the following intuition. If a position $D$ is acceptable at the date $t+1$ irrespectively of actual scenario $\omega \in \Omega$, then $D$ should also be accepted at the earlier time $t$ if we neglect the payment at date $t+1$. This payment is not taken into consideration in the definition of consistency, because it does never enter the risk evaluation at time $t+1$ by the axiom of independence of the past. In an analogous manner, *rejection consistency* states the idea that a position should already be rejected at time $t$ if we neglect the payment at $t+1$ and the position is rejected at the later date $t+1$ in any scenario $\omega \in \Omega$.

The consistency conditions have implications for the representation of a distribution-invariant dynamic risk measure given by

$$\rho_t(D) = P_t^T \cdot \Theta_t\left[\mathcal{T}_t(D)\right].$$

Let $\mathcal{N}_t \subseteq \mathcal{M}_{1,c}(\mathbb{R})$ be the acceptance set of the static risk measure $\Theta_t$. Assume that the probability space is rich in the sense that there exists a unif(0,1)-distributed random variable independent of $\mathcal{F}_{T-1}$. Then the following holds (see the proof of Corollary 2.4.2):

- If $\rho$ is acceptance consistent, then $\mathcal{N}_{t+1} \subseteq \mathcal{N}_t$.
- If $\rho$ is rejection consistent, then $\mathcal{N}_{t+1} \supseteq \mathcal{N}_t$.

If both consistency conditions are satisfied, we obtain the following corollary.

***Corollary 2.4.2.*** *Assume that the probability space is rich in the sense that there exists a unif(0,1)-distributed random variable independent of $\mathcal{F}_{T-1}$. Let the M-invariant dynamic risk measure $\rho$ be both acceptance and rejection consistent. Then $\rho$ can be represented by*

$$\rho_t(D) = P_t^T \cdot \Theta\left[\mathcal{T}_t(D)\right]$$

*Here, $\Theta$ is a unique static risk measure considered as a functional on probability measures on $\mathbb{R}$ with acceptance set*

$$\mathcal{N} = \{\mu \in \mathcal{M}_{1,c}(\mathbb{R}) : H_t(\mu) = 1\} \qquad (t = 0, 1, \ldots, T-1).$$

*Proof.* Assume that $\rho$ is acceptance consistent. Let $\mu \in \mathcal{N}_{t+1}$. Since the probability space is rich, there exists a random variable $Z \sim unif(0,1)$ independent of $\mathcal{F}_{T-1}$. We define $K = q_\mu(Z)$ where $q_\mu$ is the quantile function of $\mu$. Observe that $\mathcal{L}(K|\mathcal{F}_t) = \mathcal{L}(K|\mathcal{F}_{t+1}) = \mu$. Let $D := K \cdot e_T$. We obtain that

$$1 = H_{t+1}(\mu) = a_{t+1}(D) = a_t(D) = H_t(\mu).$$

Hence, $\mu \in \mathcal{N}_t$.

If $\rho$ is rejection consistent, the proof is analogous. $\square$

### 2.4.2 Consistency and Mixtures of Distributions

According to Corollary 2.4.2 a dynamic risk measure can be represented by one universal static risk measure, if it is both acceptance and rejection consistent. In the following theorem we take the opposite point of view asking the question:
If a dynamic risk measure can be represented by a single static risk measure - what are the properties of the static risk measure, in case the dynamic risk measure satisfies consistency properties?

It turns out that this question can be answered employing the notion of mixtures of probability measures. The following definition introduces the appropriate concept, cf. Winkler (1985).

***Definition 2.4.3.*** *Let $\mathcal{C}$ be a measurable subset of $\mathcal{M}_{1,c}(\mathbb{R})$. We say that $\mathcal{C}$ is* locally measure convex *if for all $c \in \mathbb{R}$ and any probability measure $\gamma$ on $\mathcal{C} \cap \mathcal{M}_1([-c, c])$ the mixture $\int \nu\gamma(d\nu)$ is again an element of $\mathcal{C}$.*

The last definition simply formalizes the notion of measure convex sets of probabilities in the context of measures with bounded support. The next theorem gives a first answer to our question.

***Theorem 2.4.4.*** *Let $\Theta$ be a static risk measure, and let $\mathcal{N} \subseteq \mathcal{M}_{1,c}(\mathbb{R})$ be its acceptance set. Then*

$$\rho_t(D) = P_t^T \cdot \Theta\left[\mathcal{T}_t(D)\right] \tag{2.7}$$

*defines an M-invariant dynamic risk measure. If $\mathcal{N}$ is locally measure convex, then $\rho$ is acceptance consistent. If $\mathcal{N}^c$ is locally measure convex, then $\rho$ is rejection consistent.*

*Proof.* First, $\rho$ defines a M-invariant dynamic risk measure by Lemma 2.3.11. Second, we prove that $\rho$ is acceptance consistent, if $\mathcal{N}$ is locally measure convex. The case of rejection consistency will then work analogously.
It is not difficult to see that independence of the past and invariance under adapted transforms implies that it suffices w.l.o.g. to investigate terminal positions only, i.e. positions $D \in \mathcal{D}$ of the form $D = K \cdot e_T$ with $K \in L^\infty(\Omega, \mathcal{F}, P)$. By $c \in \mathbb{R}$ we denote some real number such that $K \in [-c, c]$. Define now a kernel $Q_t$ from $(\Omega, \mathcal{F}_t)$ to $(\Omega, \mathcal{F})$ such that for measurable $A \subseteq \Omega$,

$$Q_t(\omega, A) = P(A|\mathcal{F}_t)(\omega)$$

Set $\mu_s := \mathcal{L}(K|\mathcal{F}_s)$. Then we obtain by disintegration for $P$-almost every $\omega \in \Omega$ that

$$\mu_t(\omega, \cdot) = \int \mu_{t+1}(\omega', \cdot) Q_t(\omega, d\omega')$$

Suppose that $a_{t+1}(D) \equiv 1$. Then $\mu_{t+1}(\omega', \cdot) \in \mathcal{N} \cap \mathcal{M}_1([-c,c])$ for $P$-almost all $\omega' \in \Omega$. Hence for $P$-almost all $\omega \in \Omega$,

$$\mu_t(\omega, \cdot) = \int \mu_{t+1}(\omega', \cdot) Q_t(\omega, d\omega') \in \mathcal{N},$$

since $\mathcal{N}$ is locally measure convex. This implies $a_t(D) \equiv 1$. Therefore, $\rho$ is acceptance consistent. □

The characterization of consistency in terms of the acceptance sets of the representing risk measure and mixtures of probability measures can be strengthened if the underlying probability space is rich enough.

***Theorem 2.4.5.*** *Assume that the probability space is rich in the sense that there exist both a unif(0,1)-distributed random variable independent of $\mathcal{F}_{T-1}$, and a unif(0,1)-distributed, $\mathcal{F}_{T-1}$−measurable random variable independent of $\mathcal{F}_{T-2}$. Assume again that the dynamic risk measure $\rho$ is represented as in Theorem 2.4.4.*

*Then $\rho$ is acceptance consistent, if and only if $\mathcal{N}$ is locally measure convex. Analogously, $\rho$ is rejection consistent, if and only if $\mathcal{N}^c$ is locally measure convex.*

*Proof.* We have already proven one direction in Theorem 2.4.4. Thus, we only need to show that 'consistency' implies 'measure convexity'. We will focus on the case of acceptance consistency. The case of rejection consistency works analogously.

Let $\rho$ be an M-invariant dynamic risk measure, and let $\mathcal{N}$ be the corresponding acceptance set of the representing static risk measure. Observe that $\mathcal{N}$ is measurable by definition of the functions $H_t$. Let $c \in \mathbb{R}$ be given, and let $\gamma$ be a probability measure on $\mathcal{N} \cap \mathcal{M}([-c,c])$. Let $Z \sim unif(0,1)$ be a random variable independent of $\mathcal{F}_{T-1}$, and let $U \sim unif(0,1)$ be a $\mathcal{F}_{T-1}$- measurable random variable independent of $\mathcal{F}_{T-2}$. By Borel's theorem (see Theorem 2.19 in Kallenberg (1997)) there exists a measurable function $\mu : [0,1] \to \mathcal{N}$ such that $\mu(U) \sim \gamma$. We define a kernel from $\mathcal{M}_1(\mathbb{R})$ to $\mathbb{R}$ by

$$\begin{cases} \mathcal{M}_1(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) & \to & [0,1] \\ (\nu, A) & \mapsto & \nu(A) \end{cases}$$

By the kernel randomization lemma (see Lemma 2.22 in Kallenberg (1997)) there exists a measurable function

$$q : \mathcal{M}_1(\mathbb{R}) \times [0,1] \to \mathbb{R}$$

such that $q_\nu(Z) = q(\nu, Z) \sim \nu$. Clearly, the composite function $q_{\mu(\cdot)}(\cdot) : [0,1]^2 \to \mathbb{R}$ is measurable. We define the random variable $K := q_{\mu(U)}(Z) \in [-c,c]$, and the financial position $D := K \cdot e_T \in \mathcal{D}$. We obtain that for $P$-almost all $\omega \in \Omega$,

$$\mathcal{L}(K|\mathcal{F}_{T-1})(\omega) = \mu(U(\omega)) \in \mathcal{N}, \tag{2.8}$$

$$\mathcal{L}(K|\mathcal{F}_{T-2}) = \mathcal{L}(K) = \int_{\mathcal{N}} \nu\gamma(d\nu). \tag{2.9}$$

Equation (2.8) implies $a_{T-1}(D) \equiv 1$. From acceptance consistence follows $a_{T-2}(D) \equiv 1$. Thus, $\int_{\mathcal{N}} \nu\gamma(d\nu) \overset{(2.9)}{=} \mathcal{L}(K|\mathcal{F}_{T-2}) \in \mathcal{N}$. □

### 2.4.3 Examples

Theorem 2.4.4 and Theorem 2.4.5 are very useful when constructing consistent dynamic risk measures. Examples for static risk measures which induce an acceptance and rejection consistent dynamic risk measure include the negative expected value, the worst-case measure, value at risk, and shortfall risk.

***Example 2.4.6 (Negative expected value, Worst-case measure).***
*The* negative expected value *is given by*

$$\Theta(\mu) = -\int_{\mathbb{R}} x\mu(dx).$$

*The* worst-case measure *is defined as*

$$\Theta(\mu) = -\inf\{y \in \mathbb{R} : \ \mu(-\infty, y) > 0\}.$$

*In both cases, the following holds: First, the acceptance set* $\mathcal{N} = \{\mu \in \mathcal{M}_{1,c}(\mathbb{R}) : \Theta(\mu) \leq 0\}$ *and the rejection set* $\mathcal{N}^c$ *are locally measure convex. Hence,* $\Theta$ *induces an acceptance and rejection consistent dynamic risk measure* $\rho$*, as defined in (2.7). Second,* $\Theta$ *is a coherent risk measure. Thus, the components of the corresponding dynamic risk measure* $\rho$ *are coherent on* $\mathcal{D}$*, that is for* $t = 0, 1, \dots, T-1$ *the components satisfy both convexity and positive homogeneity:*

- *Convexity:*

$$\rho_t(\alpha D + (1-\alpha)G) \ \leq \ \alpha\rho_t(D) + (1-\alpha)\rho_t(G)$$
$$(\alpha \in L^{\infty}(\Omega, \mathcal{F}_t, P), \ 0 < \alpha < 1, \ \ D, G \in \mathcal{D}).$$

- *Positive homogeneity:*

$$\rho_t(\lambda \cdot D) \ = \ \lambda \cdot \rho_t(D) \qquad (\lambda \in L^{\infty}(\Omega, \mathcal{F}_t, P), \ \lambda \geq 0, \ \ D \in \mathcal{D}).$$

***Example 2.4.7 (Value at risk).***
Value at risk *at level* $\alpha \in [0, 1)$ *is defined as*

$$\begin{aligned} \Theta(\mu) &= -\inf\{y \in \mathbb{R} : \ \mu(-\infty, y] > \alpha\} \\ &= -\sup\{y \in \mathbb{R} : \ \mu(-\infty, y) \leq \alpha\} \\ &= \ \inf\{y \in \mathbb{R} : \ \mu(-\infty, -y) \leq \alpha\}. \end{aligned}$$

*The acceptance set* $\mathcal{N} = \{\mu \in \mathcal{M}_{1,c}(\mathbb{R}) : \ \mu(-\infty, 0) \leq \alpha\}$ *and the rejection set* $\mathcal{N}^c$ *are locally measure convex. Hence,* $\Theta$ *induces an acceptance and rejection consistent dynamic risk measure* $\rho$*, as defined in (2.7).* $\Theta$ *is not a convex risk measure. Thus, the time components of the corresponding dynamic risk measure* $\rho$ *are not convex on* $\mathcal{D}$*.*

***Example 2.4.8 (Shortfall risk).***
*Let* $\ell : \mathbb{R} \to \mathbb{R}$ *be a convex loss function, i.e. an increasing, non constant and convex function. Assume that* $z$ *is an interior point of the range of* $\ell$*.*
*We define an acceptance set*

$$\mathcal{N} = \left\{\mu \in \mathcal{M}_{1,c}(\mathbb{R}) : \ \int \ell(-x)\mu(dx) \leq z\right\}.$$

$\mathcal{N}$ *induces the shortfall risk measure* $\Theta$ *by*

$$\Theta(\mu) = \inf\{m \in \mathbb{R} : \ T_m\mu \in \mathcal{N}\}.$$

*Here, for* $m \in \mathbb{R}$ *the translation operator* $T_m$ *is given by*

$$(T_m\mu)(\cdot) = \mu(\cdot - m).$$

*The induced dynamic risk measure will be denoted by* $\rho$*.*
*Shortfall risk has the following properties:*

*(1) Acceptance and rejection set are locally measure convex. Hence,* $\rho$ *is acceptance and rejection consistent.*

*(2)* $\Theta$ *is convex. Thus, the components of* $\rho$ *are convex on* $\mathcal{D}$*.*

*An exponential loss function*

$$\ell(x) = \exp(ax) \qquad (a > 0)$$

*leads to the special case of the* entropic risk measure

$$\Theta(\mu) = \frac{1}{a}\left(\log \int \exp(-ax)\mu(dx) - \log z\right).$$

## 2.5 Consistency, Compound Lotteries, and Shortfall Risk

The static risk measures representing dynamically consistent risk measures are closely related to shortfall risk. Theorem 2.5.3 will demonstrate the close link which relies on a weak closure property of the acceptance set. Before stating the theorem we need to introduce topologies on $\mathcal{M}_{1,c}(\mathbb{R})$ that allow us to deal with integrals against unbounded test functions.

For a fixed continuous function

$$\psi : \mathbb{R} \to [1, \infty)$$

we denote by $C^\psi$ the vector space of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ for which we can find a constant $c \in \mathbb{R}$ such that for all $x \in \mathbb{R}$,

$$|f(x)| \leq c \cdot \psi(x).$$

$\psi$ is called a *gauge function.* $\mathcal{M}_c^+(\mathbb{R})$ designates the space of finite measures with compact support.

***Definition 2.5.1.*** *The $\psi$-weak topology on the set $\mathcal{M}_c^+(\mathbb{R})$ is the initial topology of the family $\mu \mapsto \int f(x)\mu(dx)$ $(\mu \in \mathcal{M}_c^+(\mathbb{R}),\ f \in C^\psi)$.*

In other words, the $\psi$-weak topology is the weakest topology on $\mathcal{M}_c^+(\mathbb{R})$ for which all mappings $\mu \mapsto \int f(x)\mu(dx)$ $(\mu \in \mathcal{M}_c(\mathbb{R}))$ with $f \in C^\psi$ are continuous. It is *finer* than the weak topology. Convergence of sequences of measures can be characterized as follows:

***Lemma 2.5.2.*** *A sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{M}_c^+(\mathbb{R})$ converges $\psi$-weakly to $\mu \in \mathcal{M}_c^+(\mathbb{R})$ if and only if*

$$\int f d\mu_n \longrightarrow \int f d\mu$$

*for every measurable function $f$ which is $\mu$-almost everywhere continuous and for which exists a constant $c \in \mathbb{R}$ such that $|f| \leq c \cdot \psi$ $\mu$-almost everywhere.*

### 2.5.1 Static Risk Measures

After these preparations we are now able to state the theorem which links shortfall risk and static risk measures representing consistent dynamic risk measures. Recall that a loss function is a non decreasing function which is not identically constant.

***Theorem 2.5.3.*** *Let $\Theta$ be a risk measure on $\mathcal{M}_{1,c}(\mathbb{R})$. Assume that there exists $x \in \mathbb{R}$ with $\delta_x \in \mathcal{N}$ such that for $y \in \mathbb{R}$, $\delta_y \in \mathcal{N}^c$,*

$$(1-\alpha)\delta_x + \alpha\delta_y \in \mathcal{N} \tag{2.10}$$

*for sufficiently small $\alpha > 0$. Then the following statements are equivalent:*

*(1) Both the acceptance set $\mathcal{N}$ and the rejection set $\mathcal{N}^c$ of $\Theta$ are convex, and $\mathcal{N}$ is $\psi$-weakly closed for some gauge function $\psi : \mathbb{R} \to [1, \infty)$.*

*(2) There exists a left-continuous loss function $\ell : \mathbb{R} \to \mathbb{R}$ and a scalar $z \in \mathbb{R}$ in the interior of the convex hull of the range of $\ell$ such that*

$$\mathcal{N} = \left\{ \mu \in \mathcal{M}_{1,c}(\mathbb{R}) : \int \ell(-x)\mu(dx) \leq z \right\}.$$

*Proof.* (1) $\Rightarrow$ (2): Choose $x_1 \in \mathbb{R}$ with $\delta_{x_1} \in \mathcal{N}$ according to (2.10), and let $x_2 \in \mathbb{R}$, $\delta_{x_2} \in \mathcal{N}^c$. Define the function $g : \mathbb{R} \to \mathbb{R}$ as follows: We set $g(x_1) = 0$ and $g(x_2) = 1$. Let $z := \sup\{0 \leq \alpha \leq 1 : \alpha\delta_{x_2} + (1-\alpha)\delta_{x_1} \in \mathcal{N}\}$. Since $\mathcal{N}$ is $\psi$-weakly closed, the supremum is actually a maximum. Thus, $z \neq 1$, since $\delta_{x_2} \notin \mathcal{N}$. By (2.10) $z > 0$, hence $z \in (0,1)$. Hence, $z$ is in the interior of the convex hull of the range of $g$.

Since $\mathcal{N}$ is $\psi$-weakly closed, it follows from inverse monotonicity that there exists $r \in \mathbb{R}$ such that $[r, \infty) = \{y \in \mathbb{R} : \ \delta_y \in \mathcal{N}\}$, $(-\infty, r) = \{y \in \mathbb{R} : \ \delta_y \in \mathcal{N}^c\}$.

If $y \in [r, \infty)$, define

$$\alpha(y) := \sup\{0 \leq \alpha \leq 1 : \alpha\delta_{x_2} + (1-\alpha)\delta_y \in \mathcal{N}\}.$$

Since $\mathcal{N}$ is $\psi$-weakly closed, the supremum is actually a maximum. Thus $\alpha(y) \neq 1$, since $\delta_{x_2} \notin \mathcal{N}$. Hence, $1 - \alpha(y) \neq 0$, and we may define

$$g(y) := \frac{z - \alpha(y)}{1 - \alpha(y)}.$$

Inverse monotonicity implies additionally that $y \mapsto \alpha(y)$ is increasing on $[r, \infty)$. Hence, $y \mapsto g(y) = 1 + \frac{z-1}{1-\alpha(y)}$ is decreasing on $[r, \infty)$, since $z - 1 < 0$.

If $y \in (-\infty, r)$, define

$$\alpha(y) := \sup\{0 \leq \alpha \leq 1 : \alpha\delta_y + (1-\alpha)\delta_{x_1} \in \mathcal{N}\}.$$

Observe that $\alpha(y) \neq 1$, since $\delta_y \notin \mathcal{N}$. By (2.10) we have $\alpha(y) \neq 0$. We let

$$g(y) := \frac{z}{\alpha(y)}.$$

Inverse monotonicity implies that $y \mapsto \alpha(y)$ is increasing on $(-\infty, r)$. Hence $y \mapsto g(y)$ is decreasing on $(-\infty, r)$.

Moreover, note that on the one hand $g(y) \geq z$ for $y \in (-\infty, r)$. On the other hand, $g(y) = z + (z-1)\frac{\alpha(y)}{1-\alpha(y)} \leq z$ for $y \in [r, \infty)$, since $z - 1 < 0$. Hence, $g : \mathbb{R} \to \mathbb{R}$ is a decreasing function. We set $\ell(-x) = g(x)$. For simple probability measures $\mu \in \mathcal{M}_{1,c}(\mathbb{R})$ of the form

$$\mu = \sum_{i=1}^{n} \alpha_i \cdot \delta_{x_i},$$

$\alpha_i \geq 0$, $x_i \in \mathbb{R}$ $(i = 1, 2, \dots, n)$, $\sum_{i=1}^n \alpha_i = 1$, $n \in \mathbb{N}$, we will show that

$$\mu \in \mathcal{N} \Leftrightarrow \int g(x)\mu(dx) \leq z.$$

Let $\mu = \sum_{i=1}^n \alpha_i \cdot \delta_{x_i}$ be given. We denote by $\mathcal{M}$ the convex hull of $\{\delta_{x_i} : \ i = 1, 2, \dots, n\}$. The simplex $\mathcal{M}$ is a convex subset of the $n$-dimensional vector space spanned by $\{\delta_{x_i} : i = 1, 2, \dots, n\}$. Let $\mathcal{A} := \mathcal{N} \cap \mathcal{M}$, $\mathcal{B} = \mathcal{N}^c \cap \mathcal{M}$. Then $\mathcal{M} = \mathcal{A} \cup \mathcal{B}$, $\mathcal{A} \cap \mathcal{B} = \emptyset$, the sets $\mathcal{A}$ and $\mathcal{B}$ are both convex, and $\mathcal{A}$ is closed in the Euclidian topology. We can therefore find an affine functional $h : \mathcal{M} \to \mathbb{R}$ and $q \in \mathbb{R}$ such that

$$\begin{aligned} h(\mu) &\leq q, \qquad \mu \in \mathcal{A}, \\ h(\mu) &> q, \qquad \mu \in \mathcal{B}. \end{aligned}$$

We define

$$k := \frac{h - h(\delta_{x_1})}{h(\delta_{x_2}) - h(\delta_{x_1})}.$$

Then

$$\begin{aligned} k(\mu) &\leq \frac{q - h(\delta_{x_1})}{h(\delta_{x_2}) - h(\delta_{x_1})}, \qquad \mu \in \mathcal{A}, \\ k(\mu) &> \frac{q - h(\delta_{x_1})}{h(\delta_{x_2}) - h(\delta_{x_1})}, \qquad \mu \in \mathcal{B}. \end{aligned}$$

We show now that $g(x_i) = k(\delta_{x_i})$. For $i = 1, 2$ the claim is immediate from the definition of $k$. This implies that

$$k(\alpha\delta_{x_2} + (1 - \alpha)\delta_{x_1}) = \alpha.$$

Hence,

$$\begin{aligned} z &= \sup\{0 \leq \alpha \leq 1 : \ \alpha\delta_{x_2} + (1 - \alpha)\delta_{x_2} \in \mathcal{N}\} \\ &= \sup\left\{0 \leq \alpha \leq 1 : \ \alpha \leq \frac{q - h(\delta_{x_1})}{h(\delta_{x_2}) - h(\delta_{x_1})}\right\} = \frac{q - h(\delta_{x_1})}{h(\delta_{x_2}) - h(\delta_{x_1})}. \end{aligned}$$

Let now $i \neq 1, 2$. Assume first that $x_i \in [r, \infty)$. This implies that

$$\begin{aligned} \alpha(x_i) &= \sup\{0 \leq \alpha \leq 1 : \ \alpha\delta_{x_2} + (1 - \alpha)\delta_{x_i} \in \mathcal{N}\} \\ &= \sup\{0 \leq \alpha \leq 1 : \ \alpha + (1 - \alpha)k(\delta_{x_i}) \leq z\}. \end{aligned}$$

Observe that $\alpha(x_i) \neq 1$ and that $\alpha \mapsto \alpha + (1 - \alpha)k(\delta_{x_i})$ is continuous. Hence, the last equation is satisfied, if and only if $\alpha(x_i) + (1 - \alpha(x_i))k(\delta_{x_i}) = z$, i.e.

$$k(\delta_{x_i}) = \frac{z - \alpha(x_i)}{1 - \alpha(x_i)} = g(x_i).$$

Second, consider the case $x_i \in (-\infty, r)$. Then

$$\begin{aligned}\alpha(x_i) &= \sup\{0 \le \alpha \le 1 : \ \alpha\delta_{x_i} + (1-\alpha)\delta_{x_1} \in \mathcal{N}\} \\ &= \sup\{0 \le \alpha \le 1 : \ \alpha k(\delta_{x_i}) \ \le \ z\}.\end{aligned}$$

Observe that $\alpha(x_i) \neq 1$ and that $\alpha \mapsto \alpha k(\delta_{x_i})$ is continuous. Hence, the last equation is satisfied, if and only if $\alpha(x_i)k(\delta_{x_i}) = z$, i.e.

$$k(\delta_{x_i}) = \frac{z}{\alpha(x_i)} = g(x_i).$$

Finally, we obtain for $\mu = \sum_{i=1}^n \alpha_i \delta_{x_i}$ that

$$\mu \in \mathcal{N} \ \Leftrightarrow \ k(\mu) \le z \ \Leftrightarrow \ \sum_{i=1}^n \alpha_i g(x_i) \le z \ \Leftrightarrow \ \int g(x)\mu(dx) \le z.$$

Next we prove that $g$ is right-continuous, thus $\ell$ left-continuous. Since $g$ is decreasing, $g(x+)$ exists for each $x \in \mathbb{R}$. We have already shown that $g(x_1) < z$, $g(x_2) > z$. This implies that for given $x \in \mathbb{R}$ we can find $\alpha \in (0,1]$ and $w \in \mathbb{R}$ such that

$$\alpha g(x+) + (1-\alpha)g(w) = z.$$

Let $x_n \searrow x$. Since $g$ is decreasing, we obtain $\alpha\delta_{x_n} + (1-\alpha)\delta_w \in \mathcal{N}$ $(n \in \mathbb{N})$. Moreover, $\alpha\delta_{x_n} + (1-\alpha)\delta_w$ converges $\psi$-weakly to $\alpha\delta_x + (1-\alpha)\delta_w$. It follows that $\alpha\delta_x + (1-\alpha)\delta_w \in \mathcal{N}$, since $\mathcal{N}$ is $\psi$-weakly closed. Thus,

$$z \ \ge \ \alpha g(x) + (1-\alpha)g(w) \ \ge \ \alpha g(x+) + (1-\alpha)g(w) \ = \ z.$$

Therefore, $g(x) = g(x+)$.

Finally, we will show that the representation of $\mathcal{N}$ via the function $g$ is not restricted to simple probability measures. Let $\mu \in \mathcal{N}$. Then there exists a decreasing sequence of simple probability measures $(\mu_n)_n \subseteq \mathcal{M}_{1,c}(\mathbb{R})$ converging to $\mu$ $\psi$-weakly from above. By inverse monotonicity, $(\mu_n)_n \subseteq \mathcal{N}$, thus

$$z \ \ge \ \int g(x)\mu_n(dx) \ \to \ \int g(x)\mu(dx).$$

The convergence of the integrals follows from the right-continuity of $g$. This fact can easily be proven using Skorohod representation and Lebesgue's dominated convergence, since $g$ is bounded on a superset of the supports of the measures $(\mu_n)_n$ and $\mu$. Conversely, let $z \ge \int g(x)\mu(dx)$. Then there exists a decreasing sequence of simple probability measures $(\mu_n)_n \ \subseteq \ \mathcal{M}_{1,c}(\mathbb{R})$ converging $\psi$-weakly to $\mu$ from above. Since $g$ is decreasing, we obtain $z \geq \int g(x)\mu_n(dx)$, thus $(\mu_n)_n \subseteq \mathcal{N}$. Since $\mathcal{N}$ is $\psi$-weakly closed, we obtain $\mu \in \mathcal{N}$.

(2) $\Rightarrow$ (1): The convexity of the acceptance and rejection sets of $\Theta$ is immediate. We need to show that the acceptance set is $\psi$-weakly closed.

Let $\psi \in C(\mathbb{R})$, $\psi \geq |g|+1$ with $g(x) = \ell(-x)$ $(x \in \mathbb{R})$. We show that the functional $\mu \mapsto \int g(x)\mu(dx)$ is lower semicontinuous with respect to the $\psi$-weak topology. Since the $\psi$-weak topology on $\mathcal{M}_{1,c}(\mathbb{R})$ is metrizable, we employ the sequential characterization of closed sets. Let $z \in \mathbb{R}$ be given, and let $(\mu_n)_n \subseteq \mathcal{M}_{1,c}(\mathbb{R})$, $\mu_n \to \mu \in \mathcal{M}_{1,c}(\mathbb{R})$ $\psi$-weakly, where $\int g(x)\mu_n(dx) \leq z$ for $n \in \mathbb{N}$.

By Skorohod representation we can find bounded random variables $(X_n)_n$, $X$ on some probability space $(\Omega, \mathcal{F}, P)$ such that $X_n \sim \mu_n$ $(n \in \mathbb{N})$, $X \sim \mu$, $X_n \to X$ $P$-a.s.

We have $\lim \psi(X_n) = \psi(X)$ $P$-almost surely, and $\lim \int \psi(X_n)dP = \int \psi(X)dP$. Observe that $\psi(X_n) + g(X_n) \geq 0$ $(n \in \mathbb{N})$. By Fatou's Lemma we obtain that

$$
\begin{aligned}
\int \psi(X)dP + z \;&\geq\; \int \psi(X)dP + \liminf_n \int g(X_n)dP \\
=\quad \liminf_n \int (\psi(X_n) + g(X_n))dP \;&\geq\; \int \liminf_n (\psi(X_n) + g(X_n))dP \\
=\quad \int \psi(X)dP + \int \liminf_n g(X_n)dP \;&\geq\; \int \psi(X)dP + \int g(X)dP.
\end{aligned}
$$

The last inequality follows from the fact that $g$ is decreasing and right-continuous, since $X_n \to X$ P-almost surely. Hence,

$$
z \;\geq\; \int g(X)dP \;=\; \int g(x)\mu(dx).
$$

□

The convexity of the acceptance and rejection sets has a natural interpretation in the context of static financial positions. If two probability measures $\mu$ and $\nu$ are acceptable (resp. rejected), than for $\alpha \in [0,1]$ the compound lottery $\alpha\mu + (1-\alpha)\nu$, that randomizes over $\mu$ and $\nu$, is also acceptable (resp. rejected).

***Remark 2.5.4.***

*The risk measures characterized in the last theorem are closely connected to classical utility theory of von Neumann and Morgenstern. Setting $u(x) := -\ell(-x)$, we can interpret $u$ as a Bernoulli utility function. Then, a financial position $\mu \in \mathcal{M}_{1,c}(\mathbb{R})$ is considered acceptable, if its expected utility is larger than $-z$,*

$$
U(\mu) := \int u(x)\mu(dx) \geq -z.
$$

***Remark 2.5.5.*** *The functional $\mu \mapsto \int \ell(-x)\mu(dx)$ is $\psi$-weakly continuous for some gauge function $\psi$, if and only if $\ell$ is continuous. This follows from the representation of the dual space of $\mathcal{M}_{1,c}(\mathbb{R})$ endowed with the $\psi$-weak topology, cf. Lemma 2.5.6. Let $\psi \in C(\mathbb{R})$, $\psi \geq |g| + 1$ with $g(x) = \ell(-x)$ $(x \in \mathbb{R})$. In general, the functional is only lower semicontinuous for the $\psi$-weak topology (see the proof of Theorem 2.5.3).*

***Lemma 2.5.6.*** *Let $I : \mathcal{M}_{1,c}(\mathbb{R}) \to \mathbb{R}$ be an affine, $\psi$-weakly continuous functional. Then there exists $g \in C^\psi$ such that*

$$I(\mu) = \int g(x)\mu(dx) \qquad\qquad (\mu \in \mathcal{M}_{1,c}(\mathbb{R})).$$

*Proof.* Define $g(x) = I(\delta_x)$. If $x_n \to x$, then $\delta_{x_n} \to \delta_x$ $\psi$-weakly, hence $g(x_n) \to g(x)$. This implies that $g$ is continuous.
Suppose that $g/\psi$ is unbounded, say w.l.o.g.

$$\sup_{x \in \mathbb{R}} \frac{g(x)}{\psi(x)} = \infty.$$

Let $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ be a sequence of real numbers such that $g(x_n) \cdot (\psi(x_n))^{-1} \geq n^2$, and let

$$\mu_n = \left(1 - \frac{1}{n\psi(x_n)}\right) \cdot \delta_0 + \frac{1}{n\psi(x_n)} \cdot \delta_{x_n}.$$

Then $\mu_n \to \delta_0$ $\psi$-weakly, but

$$I(\mu_n) = \left(1 - \frac{1}{n\psi(x_n)}\right) g(0) + \frac{1}{n} \cdot \frac{g(x_n)}{\psi(x_n)}$$

diverges, a contradiction. Hence, we obtain $g \in C^\psi$.
Finally, we have to show that for $\mu \in \mathcal{M}_{1,c}(\mathbb{R})$,

$$I(\mu) = \int g(x)\mu(dx).$$

The equality does certainly hold for simple probability measures which form a dense subset of $(\mathcal{M}_{1,c}(\mathbb{R}), \tau_\psi)$. Here, $\tau_\psi$ denotes the $\psi$-weak topology. Let now $\mu \in \mathcal{M}_{1,c}(\mathbb{R})$ be arbitrary, and let $(\mu_n)_{n\in\mathbb{N}} \subseteq \mathcal{M}_{1,c}(\mathbb{R})$ be a sequence of simple probability measures converging $\psi$-weakly to $\mu$. By continuity of $I$ we get, $I(\mu_n) \to I(\mu)$. Since $g \in C^\psi$, we obtain that

$$I(\mu_n) = \int g(x)\mu_n(dx) \to \int g(x)\mu(dx).$$

$\square$

***Remark 2.5.7.*** *Condition (2.10) excludes that $\Theta$ equals the worst case measure plus some constant (say $r$), i.e.*

$$\Theta(\mu) \;=\; r \;-\; \textit{ess inf}\,\mu \qquad\qquad (\mu \in \mathcal{M}_{1,c}(\mathbb{R})).$$

***Remark 2.5.8.*** *For the* negative expected value *the loss function is given by $\ell(x) = x$ with threshold $z = 0$. For* value at risk *at level $\lambda \in (0,1)$ the loss function equals $\ell(x) = \mathbf{1}_{(0,\infty)}$ with threshold $z = \lambda$.* Shortfall risk *is already defined in terms of a loss function; characterizations and specific examples will be discussed below.*

***Remark 2.5.9.*** *For a given level $x \in [0,1)$, let $VaR_x$ be value at risk at level $x$ as defined in Example 2.4.7. For $\lambda \in (0,1)$* average value at risk *at level $\lambda$ is defined by*

$$AVaR_\lambda(\mu) = \frac{1}{\lambda}\int_0^\lambda VaR_x(\mu)dx, \quad \mu \in \mathcal{M}_{1,c}(\mathbb{R}).$$

*As the next example will show, the acceptance set of $AVaR_\lambda$ ($\lambda \in (0,1)$) is not a convex subset of the space of probability measures. Hence, $AVaR_\lambda$ does* not *satisfy condition (1) of Theorem 2.5.3, and its acceptance set cannot be represented in terms of a loss function.*

***Example 2.5.10.*** *The acceptance set of $AVaR_\lambda$ ($\lambda \in (0,1)$) is not a convex subset of the space of probability measures. For each $\lambda \in (0,1)$ this can be demonstrated by the following counterexample.*

*We let $\mu = \lambda \cdot \textit{unif}\,[-1,1] \;+\; (1-\lambda)\cdot \textit{unif}\,[1,2]$, $\nu = \delta_0$. Then we obtain for the quantile function of $\mu$ that*

$$q_\mu(\gamma) = \frac{2\gamma}{\lambda} - 1, \qquad\qquad (\gamma \le \lambda).$$

*Hence, $AVaR_\lambda(\mu) = 0$. Moreover, $AVaR_\lambda(\nu) = 0$. This implies $\mu, \nu \in \mathcal{N}$. Let $\alpha = \frac{\lambda}{2-\lambda}$. Then $q_{\alpha\nu+(1-\alpha)\mu}(\lambda) = 0$. But*

$$q_{\alpha\nu+(1-\alpha)\mu}(\gamma) = \begin{cases} \frac{2\gamma}{(1-\alpha)\lambda} - 1 \;<\; 0 & \textit{if } \gamma < \frac{(1-\alpha)\lambda}{2} \\ 0 & \textit{if } \frac{(1-\alpha)\lambda}{2} \le \gamma \le \lambda \end{cases}$$

*Hence, $AVaR_\lambda(\alpha\nu+(1-\alpha)\mu) > 0$. This implies that $\alpha\nu+(1-\alpha)\mu \notin \mathcal{N}$. The acceptance set of $AVaR_\lambda$ is therefore not a convex subset of the space of probability measures.*

The following corollary connects the preceding results with the classical theory of convex risk measures, cf. Chapter 4.6. in Föllmer and Schied (2002c).

***Corollary 2.5.11.*** *Let $\Theta$ be a risk measure on $\mathcal{M}_{1,c}(\mathbb{R})$, and assume that its acceptance set $\mathcal{N}$ is characterized as in condition (2) of Theorem 2.5.3. Then $\Theta$ is convex if and only if the loss function $\ell$ is convex.*

*Proof.* If $\ell$ is convex, the corresponding risk measure is clearly convex. We only have to prove the other direction. Assume thus that $\ell$ is not convex. Set $g(x) = \ell(-x)$. Then $g$ is not convex, and we can find $x, y \in \mathbb{R}$, $x < y$, such that

$$\frac{g(x)+g(y)}{2} \; < \; g\left(\frac{x+y}{2}\right).$$

Because $z$ is in the interior of the convex hull of the range of $g$, we can always find $w \in \mathbb{R}$ and $\alpha \in [0,1)$ such that

$$\alpha g(w) \; + \; (1-\alpha)\cdot\left(\frac{g(x)+g(y)}{2}\right) \; \leq \; z \; < \; \alpha g(w) \; + \; (1-\alpha)\cdot g\left(\frac{x+y}{2}\right).$$

We define the following random variables on $(0,1)$ with Lebesgue measure:

$$\begin{aligned} Z_1 &= w\cdot \mathbf{1}_{(0,\alpha)} \; + \; x\cdot \mathbf{1}_{[\alpha,(1+\alpha)/2)} \; + \; y\cdot \mathbf{1}_{[(1+\alpha)/2,1)} \\ Z_2 &= w\cdot \mathbf{1}_{(0,\alpha)} \; + \; y\cdot \mathbf{1}_{[\alpha,(1+\alpha)/2)} \; + \; x\cdot \mathbf{1}_{[(1+\alpha)/2,1)} \end{aligned}$$

Then $Z_1$ and $Z_2$ are both acceptable, since for $i = 1, 2$,

$$\int g(Z_i)d\lambda \; = \; \alpha g(w) \; + \; (1-\alpha)\left(\frac{g(x)+g(y)}{2}\right) \; \leq \; z.$$

We define $Z := \frac{Z_1+Z_2}{2} = w\cdot\mathbf{1}_{(0,\alpha)} + \frac{x+y}{2}\cdot\mathbf{1}_{[\alpha,1)}$, and obtain

$$\int g(Z)d\lambda \; = \; \alpha g(w) \; + \; (1-\alpha)\cdot g\left(\frac{x+y}{2}\right) \; > \; z.$$

Hence, $Z$ is not acceptable, contradicting the convexity of $\Theta$. □

Theorem 2.5.3 and Corollary 2.5.11 imply that any convex risk measure $\Theta$ on $\mathcal{M}_{1,c}(\mathbb{R})$ with locally measure convex acceptance and rejection set can be represented as *shortfall risk*, if the acceptance set is $\psi$-weakly closed for some gauge function. Shortfall risk allows a robust representation in terms of the Fenchel-Legendre transform of the associated loss function.

**Lemma 2.5.12.** *Let $\Theta$ be shortfall risk as defined in Example 2.4.8 associated with a convex and continuous loss function $\ell$. We denote the Fenchel-Legendre transform of $\ell$ by*

$$\ell^*(y) := \sup_{x\in\mathbb{R}}(yx - \ell(x)).$$

*A robust representation of the risk measure is given by*

$$\Theta(\mu) = \max_{\nu\in\mathcal{M}_1(\mu)}\left(-\int x\nu(dx) - \alpha(\nu|\mu)\right) \qquad\qquad (\mu \in \mathcal{M}_{1,c}(\mathbb{R})).$$

*Here, $\mathcal{M}_1(\mu)$ is the set of probability measures which are absolutely continuous with respect to $\mu$. The penalty function $\alpha$ is given by*

$$\alpha(\nu|\mu) = \inf_{\lambda>0} \frac{1}{\lambda}\left(z + \int \ell^*\left(\lambda\frac{d\nu}{d\mu}\right) d\mu\right) \qquad (\nu \in \mathcal{M}_1(\mu)).$$

*Proof.* We apply Theorem 4.61 of Föllmer and Schied (2002c). For $\mu \in \mathcal{M}_{1,c}(\mathbb{R})$, let $P := \mu$ and $X := id$. By $\mathcal{X}$ we denote the class of all bounded measurable functions. Of course, $(\mathbb{R}, \mathcal{B}, P)$ is not necessarily atomless. Nevertheless, if $\mathcal{L}(Y)$ $(Y \in \mathcal{X})$ denotes the distribution of $Y$ under $P$, then $\rho(Y) := \Theta(\mathcal{L}(Y))$ $(Y \in \mathcal{X})$ defines a convex risk measure on $\mathcal{X}$ which satisfies the conditions of Proposition 4.59 and Theorem 4.61 of Föllmer and Schied (2002c). This implies Lemma 2.5.12. □

***Example 2.5.13.*** *The special choice of the loss function $\ell(x) = \exp(\alpha \cdot x)$ is associated with the* entropic risk measure. *In this case, a penalty function can be defined in terms of the relative entropy:*

$$\alpha(\nu|\mu) = \frac{1}{\alpha}\left(H(\nu|\mu) - \log z\right) \qquad (\nu \in \mathcal{M}_1(\mu)).$$

*Here, the relative entropy is given by*

$$H(\nu|\mu) = \begin{cases} \int \frac{d\nu}{d\mu} \log\left(\frac{d\nu}{d\mu}\right) d\mu & \text{if } \nu \ll \mu, \\ \infty & \text{else.} \end{cases}$$

***Example 2.5.14.*** *Another example that allows explicit calculations is given by the convex loss functional*

$$\ell(x) = \begin{cases} \frac{1}{p}\, x^p & \text{if } x \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

*where $p > 1$ (see e.g. Föllmer and Schied (2002c), Example 4.64). Denoting by $q = p/(p-1)$ the dual coefficient, the Legendre-Fenchel transform is calculated as*

$$\ell^*(y) = \begin{cases} \frac{1}{q}\, y^q & \text{if } y \geq 0, \\ \infty & \text{otherwise.} \end{cases}$$

*A penalty function is then given by*

$$\alpha^p(\nu|\mu) = (p \cdot z)^{1/p} \left(\int \left(\frac{d\nu}{d\mu}\right)^q d\mu\right)^{1/q} \qquad (\nu \in \mathcal{M}_1(\mu)).$$

*The case of classical expected shortfall risk $\ell(x) = x^+$ is obtained for $p \searrow 1$. A penalty function can be calculated as*

$$\alpha(\nu|\mu) \;=\; z \cdot \left\|\frac{d\nu}{d\mu}\right\| \qquad (\nu \in \mathcal{M}_1(\mu)).$$

Finally we consider the case of coherent risk measures.

**Corollary 2.5.15.** *Let $\Theta$ be a risk measure on $\mathcal{M}_{1,c}(\mathbb{R})$, and assume that its acceptance set $\mathcal{N}$ is characterized as in condition (2) of Theorem 2.5.3. Then $\Theta$ is coherent if and only if $\ell(x) = z + \alpha x^+ - \beta x^-$ for $\alpha \geq \beta > 0$.*

*Proof.* First, let $\ell(x) = z + \alpha x^+ - \beta x^-$ be given. Since $\alpha \geq \beta > 0$, the loss function $\ell$ is convex. Hence, $\ell$ induces a convex risk measure. Let $\mu \in \mathcal{N}$, and let $X \sim \mu$ be a random variable on some atomless probability space $(\Omega, \mathcal{F}, P)$. Then for $\lambda \geq 0$,

$$\int \ell(-\lambda X)dP = z + \lambda \int (\ell(-X) - z)dP = (1-\lambda)z + \lambda \int \ell(-X)dP \leq z.$$

This implies that $\mathcal{L}(\lambda X) \in \mathcal{N}$. Hence, $\Theta$ is positively homogeneous.

Conversely, let $\Theta$ be a coherent risk measure that satisfies the hypotheses. Then $\Theta$ can be represented by a continuous and convex loss function $\ell$ and a threshold level $z \in \mathbb{R}$ in the interior of the range of $\ell$. Since $\Theta$ is positively homogeneous, $\delta_y \in \mathcal{N}$ for $y \in [0, \infty)$ and $\delta_y \in \mathcal{N}^c$ for $y \in (-\infty, 0)$. This implies that $\ell(0) = z$. Subtracting $z$, we may w.l.o.g. assume that $z = 0$ and $\ell(0) = 0$. Let $g(x) := \ell(-x)$.

Suppose that there exist $x' \in \mathbb{R}$, $\lambda' \geq 0$ such that $g(\lambda' x') \neq \lambda' g(x')$. Since $g$ is convex and $g(0) = 0$, this implies that there exist $x \in \mathbb{R}$ and $\lambda > 1$ such that $g(\lambda x) > \lambda g(x)$. Since $z = 0$ lies in the interior of the range of $g$, we can find $w_1, w_2 \in \mathbb{R}$ such that $g(w_1) < 0 < g(w_2)$. Therefore there exist $w \in \mathbb{R}$ and $\alpha \in (0, 1]$ such that

$$\alpha g(x) + (1-\alpha)g(w) = 0.$$

Hence, $\alpha\delta_x + (1-\alpha)\delta_w \in \mathcal{N}$. Since $g$ is convex with $g(0) = 0$, $g(\lambda w) \geq \lambda g(w)$. Since $g(\lambda x) > \lambda g(x)$, we obtain

$$\alpha g(\lambda x) + (1-\alpha)g(\lambda w) > 0.$$

This implies that $\alpha\delta_{\lambda x} + (1-\alpha)\delta_{\lambda w} \notin \mathcal{N}$ – contradicting the assumption of coherence. Altogether we obtain that for $x \in \mathbb{R}$, $\lambda \geq 0$ it holds that $\lambda g(x) = g(\lambda x)$. This implies that $g$ is of the form

$$g(x) = \alpha x^- - \beta x^+$$

for $\alpha, \beta \in \mathbb{R}$. $\alpha, \beta \geq 0$, since $g$ is decreasing. The inequality $\alpha \geq \beta$ follows from the convexity of $g$. Finally, $\alpha, \beta > 0$, because 0 lies in the interior of the range of $g$. □

For coherent measures of risk that satisfy the assumptions of Theorem 2.5.3 a position is acceptable, if a suitable weighted average of expected gains and expected losses is sufficiently large. In particular, gains and losses can be weighted differently, and the weight of the losses is not smaller than the weight of the gains.

While the conditions given in Theorem 2.5.3 together with convexity are all highly desirable, the additional requirement of *positive homogeneity* implicit in the notion of *coherence* has frequently been criticized in the literature. It neglects the possibility that risk might grow in a nonlinear fashion, if borrowing constraints and liquidity risk are present.

### 2.5.2 Dynamic Risk Measures

The results of the last section can be applied to dynamic risk measures. Dynamic consistency, convexity and a weak closure property imply that a dynamic risk measure can be represented in terms of shortfall risk.

***Theorem 2.5.16.*** *Assume that the filtered probability space is rich in the sense that there exist both a unif(0,1)-distributed random variable independent of $\mathcal{F}_{T-1}$, and a unif(0,1)-distributed, $\mathcal{F}_{T-1}-$measurable random variable independent of $\mathcal{F}_{T-2}$. Let $\rho$ be an M-invariant dynamic risk measure. We make the following assumptions:*

*(1) $\rho$ is acceptance and rejection consistent.*

*(2) $\rho$ is convex in the sense that for $t = 0, 1, \ldots, T-1$, $\alpha \in (0,1)$, $D, G \in \mathcal{D}$,*

$$\rho_t(\alpha D + (1-\alpha)G) \leq \alpha\rho_t(D) + (1-\alpha)\rho_t(G).$$

*(3) The set $\mathcal{N} = \{\mu \in \mathcal{M}_{1,c}(\mathbb{R}) : \; H_t(\mu) = 1\}$ $(t = 0, 1, \ldots, T-1)$ is $\psi$-weakly closed for some gauge function $\psi : \mathbb{R} \to [1, \infty)$.*

*(4) Assume that there exists $x \in \mathbb{R}$ with $\delta_x \in \mathcal{N}$ such that for $y \in \mathbb{R}$, $\delta_y \in \mathcal{N}^c$,*

$$(1-\alpha)\delta_x + \alpha\delta_y \in \mathcal{N}$$

*for sufficiently small $\alpha > 0$.*

*Then there exists a continuous and convex loss function $\ell : \mathbb{R} \to \mathbb{R}$ with associated shortfall risk measure $\Theta$ on $\mathcal{M}_{1,c}(\mathbb{R})$ such that $\rho$ can be represented as*

$$\rho_t(D) = P_t^T \cdot \Theta\left[\mathcal{T}_t(D)\right]. \tag{2.11}$$

*Proof.* By Corollary 2.4.2 there exists a unique risk measure $\Theta$ such that $\rho$ can be represented according to (2.11). By Theorem 2.4.5 the acceptance set $\mathcal{N}$ and the rejection set $\mathcal{N}^c$ are locally measure convex, thus convex. Hence, $\mathcal{N}$ can be represented according to Theorem 2.5.3 for some loss function $\ell : \mathbb{R} \to \mathbb{R}$. The convexity of $\rho$ implies the convexity of $\Theta$. This implies by Corollary 2.5.11 that $\ell$ is convex and therefore continuous. Hence, $\Theta$ is the shortfall risk measure associated with the continuous and convex loss function $\ell$. □

From the point of view of an investor or regulator, distribution-invariance at the reference time $T$, convexity, and dynamic consistency are desirable properties of a dynamic risk measure. The additional requirement on $\mathcal{N}$ to be $\psi$-weakly closed for some gauge function $\psi$ is very weak and is even economically meaningful: terminal positions which can be approximated by acceptable positions in a rather fine topology are again acceptable. We argue therefore that static shortfall risk provides a good basis for the dynamic evaluation of dynamic financial positions.[1]

## 2.6 Conclusion

We discussed an axiomatic characterization of dynamic risk measures for dynamic cash flows. For the special case of terminal financial positions at a given reference date, we require that the risk measure depends on their conditional distribution only. A key insight of the chapter is that dynamic consistency and the notion of measure convex sets of probability measures are intimately related. Measure convexity can be interpreted using the concept of compound lotteries. We characterize the class of static risk measures that represent consistent dynamic risk measures. It turns out that these are closely connected to shortfall risk. Under weak additional assumptions, static convex risk measures coincide with shortfall risk if compound lotteries of acceptable respectively rejected positions are again acceptable respectively rejected. This result implies a characterization of dynamically consistent convex risk measures.

[1]In case of additional model uncertainty, an investor or regulator should consider robust versions of the results discussed in the current chapter. Such an extension is, however, a topic of future research.

# Part II

# Models of Financial Risk

# Chapter 3

# Credit Contagion and Aggregate Losses

## 3.1 Introduction

Defaults of firms are stochastically dependent. One reason is that firms' financial health is sensitive to macro-economic factors, such as energy prices, GDP growth, or interest rates. These factors are common to all firms operating in an economy. The fluctuation of factors affects firms simultaneously and induces *cyclical default dependence*. Another reason for default dependence is the existence of business ties between firms. These links often provide the channel for the propagation of economic distress from one firm to another. This is called *credit contagion*. In this chapter we propose a model for such contagion phenomena and study the contagion-induced volatility of aggregate credit losses on large portfolios of financial positions. The measurement of aggregated risk is essential for the management and regulation of financial institutions.

Borrowing and lending networks constitute a typical distress propagation channel. In the banking sector, *interbank lending* refers to banks' mutual claims. To the extent that interbank loans are neither collateralized nor insured against, the distress of a bank may trigger the subsequent distress of other banks in the lending chain. Allen and Gale (2000) propose an equilibrium model for such phenomena where different sectors of the banking system have overlapping claims on one another in order to buffer liquidity preference shocks. This arrangement is however financially fragile: depending on the degree of connectedness of the buffer system, a small liquidity preference shock in one bank can spread through the economy and cause the distress of other banks as well. In the manufacturing sector, *trade credits* link suppliers and buyers of goods through a chain of obligations. Kiyotaki and Moore (1997) study how a liquidity shock, which leads to the distress of an individual customer in the first place, can propagate through the borrowing-lending network and result in a chain reaction.

While insightful, these micro-economic models cannot quantify aggregated loss risk due to contagion. In order to derive explicitly the distribution of aggregated losses due to contagion, we propose a reduced-form contagion model. Our approach adopts the micro-economic reasoning of Allen and Gale (2000) and Kiyotaki and Moore (1997), but models local firm interaction statistically.

We consider a homogeneous economy that hosts a large number of firms that share the same individual characteristics. The business partner network is represented by a multi-dimensional lattice, whose nodes are identified with firms. The edges represent business partner relationships, i.e. borrowing or lending contracts. The dimension of the lattice measures the denseness of the business partner network.

The financial health of a firm is characterized by the liquidity available to the firm. We specify two states, "high liquidity" and "low liquidity," the latter describing a firm that is financially distressed. The initial state of a firm is random. Over time, a firm migrates between states, reflecting a dynamic business environment. A state transition is a Poisson event, whose intensity depends on the state of the firm's business partners. We suppose that a firm's transition intensity is proportional to the number of the firm's business partners that are in the opposite state. The intuition is that a financially distressed firm is likely to default on payment obligations. The more distressed partners a healthy firm has, the higher is the likelihood that the firm suffers a liquidity shortage and becomes distressed as well.

The joint evolution of firms' states over time is described by a continuous-time Markov process. This process is also known as the *voter-model* in the theory of interacting particle systems [Liggett (1985)]. We analyze the asymptotic behavior of the liquidity process. The structure of the invariant (equilibrium) distribution of liquidity states depends on the denseness of the business partner network. If the network is not dense and a firm has only a few business partners, then an individual firm is highly dependent on each of these partners. Clusters of firms in the same state are relatively stable. Their size fluctuates randomly; they grow and merge with other clusters. In the long run, all firms are in the same state. This implies a high degree of systemic risk. In their micro-economic model, Allen and Gale (2000) obtain a qualitatively very similar behavior: with a simple lending network structure, firms are financially fragile and the degree of systemic risk is high.

If the business partner network is dense and a firm has a larger number of business partners, then the equilibrium distribution of liquidity states becomes non-trivial. Random clusters of firms in the same state appear only locally and their size fluctuates. In particular, they do not merge and grow in the same way as in a less dense network, but they are more unstable and less persistent. There are again qualitative parallels to the micro-economic contagion models of Allen and Gale (2000) and Kiyotaki and Moore (1997). With a dense borrowing/lending network, firms are more robust with respect to liquidity shocks, since these are buffered through the dense network.

We investigate the structure of the equilibrium liquidity distribution in a large homogeneous economy, where firms are equal with respect to their marginal liquidity risk. In the ergodic case, the equilibrium liquidity states are governed by an extremal distribution corresponding to the fixed degree of marginal liquidity risk. In the general case, the equilibrium liquidity states are governed by a mixture of extremal distributions. The mixing distribution corresponds to the distribution of the average number of low-liquidity firms in the whole economy, which is a random quantity. It can hence be thought of as describing systematic risk. The mixing distribution, as well as the expected proportion of low-liquidity firms, is not changed through the interaction of firms. What interaction changes is, however, the dependence between firms' states. For any finite number of firms, the probability to find many firms in the same state is higher than with independent firms.

For a fixed horizon, we are interested in the distribution of aggregate losses that a financial institution suffers from positions contracted with firms subject to credit contagion. We assume that the loss on a position with a given firm is random and depends on the firm's liquidity state. Given the firms' states, losses are independent. We base our assessment of aggregate loss risk on the equilibrium liquidity distribution in a dense business partner network.

Average losses on infinitely large portfolios are determined through the average proportion of low-liquidity firms and the expected conditional position losses. While loss uncertainty stemming from the fluctuation of position losses averages out, loss uncertainty remains from the average proportion of low-liquidity firms. The randomness in average losses is hence governed by the mixing distribution, which represents systematic risk. Contagion effects play no role in infinitely large portfolios, they are diversified away entirely. This is confirmed by an analysis of the quantiles of the loss distribution, which are basically governed by the quantiles of the mixing distribution, if the portfolio is very large.

Losses in finite portfolios are our main concern. We provide an *explicit* Gaussian approximation to the distribution of losses in finite portfolios. This approximation is based on a recent central limit theorem for the general voter model proved by Zähle (2001), which is non-classical due to the strong dependence induced by the local interaction. The approximation is the key to the measurement and management of the portfolio's aggregated risk.

We analyze the determinants of the volatility of losses in finite portfolios. As in infinite portfolios, average losses are random and governed by the distribution of systematic risk. However, in finite portfolios contagion induces a second-order effect on the volatility of losses. It causes additional fluctuations of losses around their (random) averages, so that the probability of large losses is elevated through contagion. The amount of additional loss volatility depends on two quantities: the characteristics of the systematic risk in the economy and the denseness of the business partner network. Through numerical calculations we illustrate that the effect of contagion on losses decreases with increasing volatility of systematic risk and denseness of the network.

Our approximation results complement the existing literature on large homogeneous credit portfolios, which neglects credit contagion and instead focuses on cyclical default dependence: Vasicek (1987), Frey and McNeil (2002), Lucas et al. (2001), Schloegl (2002) and Gordy (2001). In these models, the volatility of aggregate losses is entirely due to the fluctuation of some exogenous macro-economic variables. Ultimately, portfolio models have to account for both cyclical dependence and credit contagion. Based on the results in this chapter, Giesecke and Weber (2004b) provide an approximation that integrates both effects. They quantify the relation between loss volatility, volatility of macro-economic factors, and denseness of the business partner network. Alternative approaches to jointly model cyclical and contagion effects include those by Jarrow and Yu (2001), Giesecke (2001), Schönbucher and Schubert (2001). These contributions do not consider portfolio losses, but focus on the joint default distribution for a set of heterogeneous firms and the pricing of associated credit sensitive securities. Our reduced form contagion model is close in spirit to the reduced form model of Jarrow and Yu (2001), who assume that a firm's default intensity depends on the default status of the firm's counterparties. A similar idea appears in Davis and Lo (2001), who consider finite portfolios. Their conclusions are qualitatively similar to ours.

The remainder of this chapter is organized as follows. In Section 2 we propose a statistical model for credit contagion. In Section 3, we analyze the asymptotic behavior of the liquidity process and the structure of the equilibrium liquidity distribution. In Section 4 we provide an explicit approximation of the distribution of aggregate losses. Section 5 concludes by discussing the model assumptions.

## 3.2 Modeling Credit Contagion

We provide a statistical model for the effects of credit contagion and investigate their consequences on the level of both the whole economy and large portfolios.

### 3.2.1 A Reduced-Form Model

We consider an economy with a collection $S$ of small or medium sized firms which is at most countably infinite. A firm $i \in S$ can be in two liquidity states, denoted 0 and 1. State 0 corresponds to high liquidity, while state 1 corresponds to low liquidity. The state of the economy is characterized by a configuration in the state space $\{0,1\}^S$. We are interested in the evolution of firms' liquidity states over time and the interdependence of the states of different firms.

The liquidity state of an arbitrary firm $i$ is influenced by the state of a collection $N(i) \subseteq S \setminus \{i\}$ of business partners. We assume that any firm $i \in S$ is a creditor of its business partners. At a time $\tau$ some business partner $j \in N(i)$ is obliged to pay a certain amount to its creditor $i$. Depending on its liquidity state at the maturity date $\tau$, firm $j$ will or will not fulfill its obligation. We suppose that firm $j$ pays its debt if it is in the high liquidity state 0. If it is in the low liquidity state 1, then it defaults on its obligation. Hence the state of obligor $j$ influences the liquidity state of the creditor firm $i$. If $j$ fulfills its obligation, creditor $i$ is in the high liquidity state 0 from time $\tau$ onwards. Otherwise, $i$ is in the low liquidity state 1 after time $\tau$.

In a large economy, modeling explicitly all borrowing and lending relationships becomes extremely complex and is not tractable. To reduce the complexity of the problem, we provide a statistical model for the interaction of the firms. In contrast to standard micro-economic models, we thus describe the choice of the obligor $j \in N(i)$ and the maturity date $\tau$ in a probabilistic way. We assume that $\tau$ is a random time that is exponentially distributed with parameter 1. The business partner of firm $i$ whose payment is due at time $\tau$ is chosen according to some distribution $p(i,j)$ where $j \in N(i)$.

We denote the liquidity state of a firm $i \in S$ by $\xi(i)$. Given our assumptions, the transition between liquidity states is a Poisson event. The transition rate of the state of firm $i \in S$ can formally be written as

$$c(i,\xi) = \sum_{j:j\in N(i)} p(i,j)|\xi(i) - \xi(j)|.$$

In this sense our credit contagion model belongs to the class of *reduced-form* credit risk models [see, e.g., Jarrow and Turnbull (1995), Duffie et al. (1996), Duffie and Singleton (1999), Jarrow et al. (1997), and Lando (1998) for single-firm models]. The idea that a firm's default intensity directly depends on the state of its counterparties has recently appeared in Jarrow and Yu (2001) and Davis and Lo (2001).

### 3.2.2 The Voter Model

The evolution of the liquidity state of an arbitrary firm $i$ is influenced by the state of a collection $N(i) \subseteq S \setminus \{i\}$ of business partners. $N(i)$ will be called the set of *neighbors* of firm $i$. For simplicity, we assume that firms influence each other in a *symmetric* way: if firm $i$'s state is influenced by firm $j$, then firm $j$'s state is influenced by firm $i$. Expressed in terms of the neighborhoods,

$$j \in N(i) \Rightarrow i \in N(j).$$

If we connect all firms $i \in S$ to their neighbors $j \in N(i)$, we get an undirected graph which characterizes the business relations of the firms. Business partners are nearest neighbors on the graph. For tractability, we assume a simple neighborhood structure which is specified by a $d$-dimensional lattice. In particular, all firms have the same finite number of business partners. Hence, we consider an economy with a countably infinite number of firms. Firms are identified with their location on the $d$-dimensional integer lattice $S = \mathbb{Z}^d$. According to our assumptions, at a unit exponential time the payment of a business partner $j$ of firm $i$ is due. Firm $j$ is chosen according to some distribution $p(i,j)$ where $|j-i| = 1$. Here $|\cdot|$ denotes the length of the shortest path between two firms on the lattice. To keep the analysis simple, we choose $p(i,j)$ to be the uniform distribution, i.e. $p(i,j) = \frac{1}{2d}$. The contagion pattern we proposed above implies that the transition rate $c$ is given by

$$c(i,\xi) \;=\; \frac{1}{2d} \sum_{j:|i-j|=1} |\xi(i) - \xi(j)| \;=\; \begin{cases} \frac{1}{2d} \sum_{j:|i-j|=1} \xi(j) & \text{if} \quad \xi(i) = 0 \\ \frac{1}{2d} \sum_{j:|i-j|=1} [1 - \xi(j)] & \text{if} \quad \xi(i) = 1. \end{cases}$$

The transition rate $c$ is a function of the firm $i \in \mathbb{Z}^d$ and the liquidity configuration $\xi \in X := \{0,1\}^{\mathbb{Z}^d}$. A regular version of the process is given by the *voter model*. The voter model is well-known in the theory of interacting particle systems [Liggett (1985), Liggett (1999)]. The evolution of firms' liquidity states is described by a continuous-time Feller process $(\eta_t)_{t \geq 0}$ with state space $X$ and transition rate $c$. Here $\eta_t(i)$ is the liquidity state of firm $i$ at time $t$.

The rate at which firm $i$ switches its state is represented by $c(i,\xi)$. That is, a firm $i \in \mathbb{Z}^d$ with a high liquidity state ($\xi(i) = 0$) migrates to a low liquidity state ($\xi(i) = 1$) at a rate proportional to the number of low-liquidity neighboring firms

$j \in \{j : \xi(j) = 1, |i-j| = 1\}$, and vice versa. Put another way, after a unit exponential waiting time in one state, a firm $i \in \mathbb{Z}^d$ migrates to the state of some neighboring firm $j \in \{j : |i-j| = 1\}$ which is chosen with probability $1/2d$. A transition is hence a Poisson event, whose intensity is proportional to the number of neighboring firms with opposite liquidity state. It is easy to see that if all firms $i \in \mathbb{Z}^d$ are either in good or in bad shape, then the transition rate is zero.

***Remark 3.2.1.*** *We could multiply the transition rate $c$ of the voter model by an arbitrary constant without changing the long-run behavior of the dynamics. The modified rate translates into a deterministic linear time change of the model.*

This formal model of the joint evolution of firms' liquidity states probabilistically describes the pattern of credit contagion phenomena as we introduced them above. Pick the specific example of trade credit. If a high-liquidity firm's business partners in a trade credit are in the low-liquidity state, then the probability that this firm migrates to the low-liquidity state due to a payment default in the credit chain increases with the number of low-liquidity partners. If a low-liquidity firm's business partners are in the high-liquidity state so that a default in the chain is unlikely, then the probability of that firm's migration to the high-liquidity state increases with the number of healthy partners.

## 3.3 Equilibrium Behavior

Let us now look at the equilibrium distributions (i.e. invariant distributions) and the asymptotic behavior of the liquidity process $\eta$. It turns out that the structure of the equilibrium distributions depends on the dimension $d$. The dimension $d$ determines the denseness of the business partner network. The larger $d$, the more business partners has any individual firm. At the same time the number of indirect inter-firm links of given length increases. More specifically, if $i$ and $j$ are two firms, a sequence $(i_0, i_1, \ldots, i_n)$ of firms is a link of length $n$ between $i$ and $j$, if $i_k$ is a neighbor of $i_{k+1}$ for $k \in \{0, 1, \ldots, n-1\}$, $i_0 = i$ and $i_n = j$. The number of links of length $n$ emanating from a given firm equals $2d(2d-1)^{n-1}$ and grows exponentially in $n$ and polynomial in $d$.

### 3.3.1 Non-Dense Business Partner Network

The liquidity state of an obligor $i$ is revealed to its creditor $j \in N(i)$ at the maturity of an obligation. Firm $i$ will default if and and only if it is in the low-liquidity state. Conversely, immediately after maturity, the liquidity state of both firms $i$ and $j$ will be equal according to the contagion process. Thus, also the subsequent state of firm $j$ will be known to firm $i$. Hence, for firms in the network the liquidity states of business partners are partially observable. For financial institutions *outside* of the network of small or medium sized firms we suppose in contrast that the liquidity state of firms $i \in \mathbb{Z}^d$ cannot be observed. Hence the liquidity configuration of the firms is random and described by a probability distribution on the state space $X$.

At some initial time the distribution of $\eta$ is given by the distribution $\mu$ on $X$. We are interested in the behavior of the liquidity process in the long run. The process $\eta$ has càdlàg paths; for convenience, we will work with the canonical version of the process. $\Omega$ denotes the space of càdlàg functions on $\mathbb{R}_+$ with values in $X$ endowed with the usual augmented filtration. For the law of the process $\eta$ with initial distribution $\mu$ we will write $P^\mu$.

***Definition 3.3.1.*** *For $\xi \in X$ and $i \in \mathbb{Z}^d$ we define the translation $T_i(\xi)(j) = \xi(i+j)$. Canonically, the translation $T_i$ operates also on subsets of $X$. A measure $\mu$ on $X$ is called translation-invariant, if $\mu(A) = \mu(T_i A)$ for all $i \in \mathbb{Z}^d$ and for all measurable $A \subseteq X$.*

We shall assume that $\mu$ is translation-invariant and denote by

$$\rho = \mu\{\xi : \xi(i) = 1\} \tag{3.1}$$

the Bernoulli parameter of the initial marginal liquidity distribution for an arbitrary firm $i$. $\rho$ can hence be thought of as a measure of an individual firm's marginal liquidity risk. In particular, the translation-invariance of $\mu$ implies that the firms in the economy are *homogeneous* with respect to marginal risk.

For $d = 1, 2$ and translation-invariant initial law $\mu$, as $t \to \infty$ the distribution of $\eta_t$ converges weakly to the mixture

$$\rho\delta_1 + (1-\rho)\delta_0, \tag{3.2}$$

cf. Liggett (1999). Here $\delta_\xi$ is the Dirac measure placing mass 1 on configuration $\xi \in X$. In (3.2) the indices 0 (1) refer to the configurations with *all* firms being in high (low) liquidity state. The liquidity process $\eta$ *clusters*, i.e. for all $i, j \in \mathbb{Z}^d$

$$\lim_{t\to\infty} P^\mu[\eta_t(i) \neq \eta_t(j)] = 0. \tag{3.3}$$

If the business partner network is not dense, then in the long run only one firm type appears: with probability $\rho$ *all* firms are in the low-liquidity state, and with probability $1-\rho$ *all* firms are in the high-liquidity state. The marginal liquidity distribution of any individual firm is invariant under the contagion dynamics: the degree of marginal risk is not affected by the interaction process. Nevertheless, the economy can change drastically on the macroscopic level.

This behavior is quite intuitive in the trade credit chain interpretation. If initially the marginal probability $\rho$ of individual firms to be in the low-liquidity state is high, then it is quite likely that high-liquidity firms in the credit chain become "infected." Random clusters of firms in the low-liquidity state emerge with high probability, while clusters of firms in the high-liquidity state emerge only with low probability. In any case, if the chain a firm operates in is "short," then the state of the relatively few business partners highly dominates the state of a firm in the chain. Here clusters of firms of the same type are relatively stable. The size of the clusters changes through random fluctuations, but for low dimensions $d \leq 2$ some of the clusters merge and form large growing clusters. Asymptotically, with high probability $\rho$ *all* firms are in the low-liquidity state, and with low probability $1-\rho$ *all* firms are in the high-liquidity state. Vice versa, if $\rho$ is low, then it is unlikely that a firm gets distressed. In the limit, with high probability $1-\rho$ *all* firms will have the high-liquidity state, with low probability $\rho$ the low-liquidity state.

### 3.3.2 Dense Business Partner Network

The limiting behavior of $\eta$ differs for higher dimensions $d > 2$. In the current section we analyze the asymptotic behavior of the liquidity process for $d > 2$ in two steps. We focus first on the special situation of *ergodic* initial distributions. Then we derive the long-run behavior of the process for general translation-invariant initial distributions from this special case using a refined Choquet decomposition.

***Definition 3.3.2.*** *A translation-invariant distribution $\mu$ on $X$ is called* ergodic*, if $\mu(A) \in \{0,1\}$ for any translation-invariant subsets $A \subseteq X$. Here, a set $A \subseteq X$ is called translation-invariant, if $T_i A = A$ for all $i \in \mathbb{Z}^d$.*

Before we characterize the asymptotic behavior of the liquidity process, we need to describe the structure of probability measures which are invariant for the voter model. We endow the space $\mathcal{M}_1(X)$ of probability measures on $X$ with the weak topology. The set of probability measures which are invariant for the voter model is denoted by $\mathcal{I}$. The collection of invariant measures $\mathcal{I}$ is a convex set which is closed in the weak topology. The set of extremal points of $\mathcal{I}$ is denoted by $\mathcal{I}_{ex}$. Here, a measure $\nu \in \mathcal{I}$ is called an *extremal point* of $\mathcal{I}$, if $\nu$ is not a proper convex combination of other elements of $\mathcal{I}$. That is, if $\nu = \alpha\nu_1 + (1-\alpha)\nu_2$ for $\nu_1, \nu_2 \in \mathcal{I}$, $\alpha \in (0,1)$, then $\nu = \nu_1 = \nu_2$.

It turns out that for $\rho \in [0,1]$ the set $\mathcal{I}_{ex}$ contains exactly one element $\nu_\rho$ with $\nu_\rho\{\xi :\ \xi(i) = 1\} = \rho$. We can therefore label the extremal invariant measures by the

Bernoulli parameter of their one-dimensional marginals and obtain in a natural way a one-parameter family

$$\mathcal{I}_{ex} = \{\nu_\rho : \ \rho \in [0,1]\}.$$

For the following result we refer to Liggett (1999).

***Theorem 3.3.3.*** *For any translation-invariant ergodic initial distribution $\mu$ with Bernoulli parameter $\rho = \mu\{\xi : \xi(i) = 1\}$, as $t \to \infty$ the distribution of $\eta_t$ converges weakly to the non-trivial* extremal *invariant measure $\nu_\rho$ of the voter model in dimension $d$ with parameter $\rho = \nu_\rho\{\xi : \xi(i) = 1\}$.*

In contrast to the case $d \leq 2$, in this case the contagion process $\eta$ *coexists*, referring to the lack of clustering of liquidity states in the long run. For a dense network of firms the contagion process is non-trivial. The average number of low-liquidity firms in the whole economy is a preserved quantity of the dynamics and equals $\rho$ forever.

We study the equilibrium distribution of liquidity states in case $d > 2$ for general, i.e. not necessarily ergodic initial distribution $\kappa$. As stated in the next theorem, the liquidity process $\eta$ converges weakly to a mixture of the extremal invariant measures $\nu_\rho$ ($\rho \in [0,1]$) of the voter model. A sufficient statistic for the asymptotic distribution of the process is given by the empirical proportion of low-liquidity firms in the whole economy. It is a standard result that this quantity exists almost surely for translation-invariant measures on $X$.

***Definition 3.3.4.*** *For a translation-invariant probability measure $\mu$ on $X$ the empirical proportion of low-liquidity firms is a random variable $\bar{\rho}$ which is $\mu$-almost surely defined as*

$$\bar{\rho} := \lim_{n\to\infty} |\Lambda_n|^{-1} \sum_{i \in \Lambda_n} \xi(i).$$

*Here*

$$\begin{aligned} \Lambda_n &:= [-n,n]^d \\ &:= \{i \in \mathbb{Z}^d : \ -n \leq i_1 \leq n, \ -n \leq i_2 \leq n, \ \ldots, \ -n \leq i_d \leq n\}. \end{aligned} \tag{3.4}$$

By $\mathcal{M}_e$ we denote the space of ergodic probability measures on $X$ endowed with weak topology. Let $\mathcal{G}$ be the Borel $\sigma$-algebra on $\mathcal{M}_e$. We write $\mathcal{M}_{e,\rho}$ for the subspace of $\mathcal{M}_e$ of probability measures $\nu$ with $\nu\{\xi : \xi(0) = 1\} = \rho \in [0,1]$. The theorem of Choquet states that any shift-invariant probability measure on $X$ can be represented as a mixture of ergodic measures [Georgii (1988), Theorem 14.10.]:

***Theorem 3.3.5.*** *Let $\mu$ be a translation invariant probability measure on $X$. Then there exists a probability measure $\hat{\gamma}$ on $\mathcal{M}_e$ such that $\mu = \int_{\mathcal{M}_e} \nu \hat{\gamma}(d\nu)$, i.e. for all continuous functions $f \in C(X)$ it holds that $\mu(f) = \int_{\mathcal{M}_e} \nu(f)\hat{\gamma}(d\nu)$.*

We need a refined Choquet decomposition which can be used to establish the complete convergence theorem for $\eta$ in case $d > 2$ for general translation invariant initial distributions of liquidity states:

***Theorem 3.3.6.*** *Let $\mu$ be a translation invariant probability measure on $X$. Then there exists a probability measure $Q$ on $[0,1]$ and a kernel*

$$\gamma_{\cdot}(\cdot) : \begin{cases} \mathcal{G} \times [0,1] & \rightarrow \quad [0,1] \\ (A, \rho) & \mapsto \quad \gamma_\rho(A) \end{cases}$$

*with $\gamma_\rho(\mathcal{M}_{e,\rho}) = 1$ such that*

$$\mu = \int_{[0,1]} \left( \int_{\mathcal{M}_e} \nu \gamma_\rho(d\nu) \right) Q(d\rho).$$

*Let $\hat{\gamma}$ be defined as in Theorem 3.3.5. $Q$ has the cumulative distribution function $G$ given by*

$$G(\rho) = \hat{\gamma}\{\nu \in \mathcal{M}_e : \nu\{\xi : \xi(0) = 1\} \leq \rho\}.$$

*Proof of Theorem 3.3.6.* For $\nu \in \mathcal{M}_e$ define $Y(\nu) := \nu\{\xi : \xi(0) = 1\}$. $Y$ is measurable, since $Y$ is continuous. Define $\mathcal{F} := \sigma(Y)$. Let

$$\hat{\gamma}(\cdot|\mathcal{F})(\cdot) : \begin{cases} \mathcal{M}_e \times \mathcal{G} & \rightarrow \quad [0,1] \\ (\nu, A) & \mapsto \quad \hat{\gamma}(A|\mathcal{F})(\nu) \end{cases}$$

be a regular version of the conditional probability, i.e. for fixed $A \in G$ the random variable $\hat{\gamma}(A|\mathcal{F})$ is $\mathcal{G}$-measurable, and for fixed $\nu \in \mathcal{M}_e$ $\hat{\gamma}(\cdot|\mathcal{F})(\nu)$ is a probability measure.

For $A \in \mathcal{G}$, $\hat{\gamma}(A|\mathcal{F})$ is $\sigma(Y)$-measurable. By Doob's functional representation theorem there exists a measurable mapping $\gamma_{\cdot}(A) : [0,1] \rightarrow [0,1]$ such that

$$\hat{\gamma}(A|\mathcal{F})(\nu) = \gamma_{Y(\nu)}(A).$$

Since $\hat{\gamma}(\cdot|\mathcal{F})(\cdot)$ is regular, it follows that $\gamma_{\cdot}(\cdot)$ is a kernel.

For $C \in \mathcal{G}$ we have

$$\begin{aligned} \hat{\gamma}(C) &= \int_{\mathcal{M}_e} \hat{\gamma}(C|\mathcal{F})(\nu)\hat{\gamma}(d\nu) = \int_{\mathcal{M}_e} \gamma_{Y(\nu)}(C)\hat{\gamma}(d\nu) \\ &= \int_{[0,1]} \gamma_\rho(C) Q(d\rho), \end{aligned}$$

where $Q = \mathcal{L}(Y;\hat{\gamma}) = \hat{\gamma} \circ Y^{-1}$.

This implies for $f \in C(X)$:

$$\mu(f) = \int_{\mathcal{M}_e} \nu(f)\hat{\gamma}(d\nu) = \int_{[0,1]} \left( \int_{\mathcal{M}_e} \nu(f)\gamma_\rho(d\nu) \right) Q(d\rho).$$

Note that for $\rho \in [0,1]$ $\mathcal{M}_{e,\rho}$ is measurable, since it is a closed set in $\mathcal{M}_e$. Then for any $\nu \in \mathcal{M}_{e,\rho}$:

$$\gamma_\rho(\mathcal{M}_{e,\rho}) = \gamma_{Y(\nu)}(\mathcal{M}_{e,\rho}) = \hat{\gamma}(\mathcal{M}_{e,\rho}|\mathcal{F})(\nu).$$

Observe now that

$$\int_{\mathcal{M}_{e,\rho}} \hat{\gamma}(\mathcal{M}_{e,\rho}|\mathcal{F})(\nu)\hat{\gamma}(d\nu) = \hat{\gamma}(\mathcal{M}_{e,\rho}) = \int_{\mathcal{M}_{e,\rho}} \hat{\gamma}(d\nu).$$

Hence, $\gamma_\rho(\mathcal{M}_{e,\rho}) = 1$.

Finally, we have to show that $Q$ has cumulative distribution function $G$:

$$Q((-\infty,\rho]) = \hat{\gamma}(Y \le \rho) = \hat{\gamma}\{\nu : \nu\{\xi : \xi(0) = 1\} \le \rho\} = G(\rho).$$

This completes the proof. □

Having proved this refined decomposition, we are now in a position to establish a complete convergence theorem:

***Theorem 3.3.7.*** *Let $d > 2$ and denote by $\mu_t^\kappa$ the distribution of $\eta_t$ for given initial distribution $\kappa$ on $X$. Let $\kappa$ be a translation invariant measure, and let*

$$\kappa = \int_{[0,1]} \left( \int_{\mathcal{M}_e} \nu\gamma_\rho(d\nu) \right) Q(d\rho) \tag{3.5}$$

*be the refined ergodic decomposition of $\kappa$, cf. Theorem 3.3.6. Then we have that*

$$\mu_t^\kappa = \int_{[0,1]} \left( \int_{\mathcal{M}_e} \mu_t^\nu \gamma_\rho(d\nu) \right) Q(d\rho), \tag{3.6}$$

*and, letting the symbol "$\xrightarrow{w}$" denote weak convergence of probability measures,*

$$\mu_t^\kappa \xrightarrow{w} \int_{[0,1]} \nu_\rho Q(d\rho), \tag{3.7}$$

*where $\nu_\rho$ is the extremal invariant measure of the basic voter model in dimension $d > 2$ with parameter $\rho \in [0,1]$.*

*$Q$ is the distribution of the empirical proportion of low-liquidity firm in the whole economy under the measure $\mu$.*

*Proof.* Let $f \in C(X)$. Writing $\mu_t^{\xi}$ instead of $\mu_t^{\delta_\xi}$, we have

$$\begin{aligned}\mu_t^{\kappa}(f) &= \int \mu_t^{\xi}(f)\kappa(d\xi) \\ &= \int_0^1 \left( \int_{\mathcal{M}_e} \left( \int \mu_t^{\xi}(f)\nu(d\xi) \right) \gamma_\rho(d\nu) \right) Q(d\rho) \\ &= \int_0^1 \left( \int_{\mathcal{M}_e} \mu_t^{\nu}(f)\gamma_\rho(d\nu) \right) Q(d\rho)\end{aligned}$$

Moreover, by the bounded convergence theorem

$$\begin{aligned}\lim_{t\to\infty} \mu_t^{\kappa}(f) &= \lim_{t\to\infty} \int_0^1 \left( \int_{\mathcal{M}_e} \mu_t^{\nu}(f)\gamma_\rho(d\nu) \right) Q(d\rho) \\ &= \int_0^1 \left( \int_{\mathcal{M}_e} (\lim_{t\to\infty} \mu_t^{\nu}(f))\gamma_\rho(d\nu) \right) Q(d\rho)\end{aligned}$$

since $|\mu_t^{\nu}(f)| \leq ||f||_\infty < \infty$. Noting that $\lim_{t\to\infty} \mu_t^{\nu}(f) = \nu_\rho(f)$ on $\mathcal{M}_{e,\rho}$ and that $\gamma_\rho(\mathcal{M}_{e,\rho}) = 1$, we have

$$\lim_{t\to\infty} \mu_t^{\kappa}(f) = \int_0^1 \nu_\rho(f)Q(d\rho),$$

which is our assertion.

□

The refined ergodic decomposition (3.5) describes the initial distribution $\kappa$ of liquidity states as a two-step random process: first the parameter $\rho \in [0,1]$ is chosen according to the distribution $Q$, which then prescribes the translation invariant regime

$$\kappa_\rho := \kappa_{\rho,0} := \int_{\mathcal{M}_e} \nu\gamma_\rho(d\nu).$$

The distribution $Q$ governs the mixture of the regimes $\kappa_\rho$ in the decomposition of the initial distribution.

The evolution of the liquidity distribution is described by (3.6) and (3.7). If the initial distribution $\kappa$ can be decomposed as in (3.5), then the liquidity distributions $\mu_t^{\kappa}$ at time $t$ and $\mu_\infty^{\kappa} = \lim_{t\to\infty} \mu_t^{\kappa}$ can be decomposed analogously. Theorem 3.3.7 describes these distributions of liquidity states as two-step random processes: first the parameter $\rho \in [0,1]$ is chosen according to the distribution $Q$, which then determines the regimes

$$\kappa_{\rho,t} = \begin{cases} \int_{\mathcal{M}_e} \mu_t^{\nu}\gamma_\rho(d\nu) & \text{if } t < \infty, \\ \nu_\rho & \text{if } t = \infty. \end{cases}$$

Asymptotically, the distribution $\mu_\infty^\kappa$ of liquidity states is a probability-weighted average of extremal invariant measures $\nu_\rho$ of the voter model; this mixture is governed by the distribution $Q$ which is given by the initial law of the average number $\bar{\rho}$ of low-liquidity firms in the economy.

***Corollary 3.3.8.*** *Under the assumptions of Theorem 3.3.7, we can characterize the behavior of the empirical proportion of low-liquidity firms $\bar{\rho}$ as follows:*

*(1) $\bar{\rho}$ is $\kappa_{\rho,t}$-almost surely equal to $\rho$ for $\rho \in [0,1]$ and $t \in [0,\infty]$.*

*(2) For $t \in [0,\infty]$ the law of $\bar{\rho}$ under $\mu_t^\kappa$ equals $Q$.*

*Proof.* Part (1) can be verified as follows. For $t = \infty$ the claim holds by a strong law of large numbers, since $\nu_\rho$ is ergodic with $\nu_\rho\{\xi : \ \xi(i) = 1\} = \rho$. Let $\kappa_{\rho,t}$ be given for $t \in [0,\infty)$. According to Theorem 3.3.6 there exists a kernel $\gamma'$ and a measure $Q'$ on $[0,1]$ such that

$$\kappa_{\rho,t} = \int_{[0,1]} \left( \int_{\mathcal{M}_e} \nu \gamma'_{\rho'}(d\nu) \right) Q'(d\rho').$$

Suppose that $\kappa_{\rho,t}(\bar{\rho} = \rho) < 1$. Suppose also that $Q' = \delta_\rho$. Thus,

$$\kappa_{\rho,t}(\bar{\rho} = \rho) = \left( \int_{\mathcal{M}_e} \nu \gamma'_\rho(d\nu) \right) (\bar{\rho} = \rho) = \int_{\mathcal{M}_{e,\rho}} \nu(\bar{\rho} = \rho) \gamma'_\rho(d\nu) = 1,$$

a contradiction. Thus, $Q' \neq \delta_\rho$. Hence by Theorem 3.3.7,

$$\nu_\rho = \lim_{t\to\infty} \kappa_{\rho,t} = \lim_{t\to\infty} \int_{[0,1]} \left( \int_{\mathcal{M}_e} \nu \gamma'_{\rho'}(d\nu) \right) Q'(d\rho') = \int_{[0,1]} \nu_{\rho'} Q'(d\rho'),$$

a contradiction. Thus, $\kappa_{\rho,t}(\bar{\rho} = \rho) = 1$.

Part (2) can be proven as follows:

$$\begin{aligned} \mu_t^\kappa(\bar{\rho} \le \rho') &= \int_{[0,1]} \int_{\mathcal{M}_e} \mu_t^\nu(\bar{\rho} \le \rho') \gamma_\rho(d\nu) Q(d\rho) \\ &= \int_{[0,1]} \mathbf{1}_{(-\infty,\rho']}(\rho) Q(d\rho) = Q(-\infty, \rho'). \end{aligned}$$

□

The distribution of the average number of low-liquidity firms in the economy is preserved under the contagion dynamics; it is not changed through the interdependence of firms. What interaction between firms changes is the dependence between the liquidity states of different firms. For any finite number of firms, the probability to find many firms in the same liquidity state is higher than in the case of independent firms.

## 3.4 Aggregate Losses on Large Portfolios

In the previous section we modeled credit contagion phenomena induced by the interdependence of firms and analyzed the weak convergence of the liquidity process $\eta$ and its distribution in equilibrium. In this section we consider the aggregate losses associated with credit contagion phenomena. Throughout, we suppose that the economy is in equilibrium, in the sense that the distribution of firms' liquidity states is invariant.

We consider a financial institution that holds a portfolio of financial positions issued by firms $i \in \Lambda_n \subseteq \mathbb{Z}^d$, where $\Lambda_n$ is defined in (3.4). The parameter $n \in \mathbb{N}$ determines the size of the portfolio $\Lambda_n$. The positions are subject to credit risk: whether or not an issuer will be able to honor a financial obligation depends on the issuer's state. We wish to assess the bank's aggregated risk of losses at some fixed horizon. Denoting the losses on positions contracted with firm $i \in \Lambda_n$ by the random variable $U(i)$, we are interested in the distribution of portfolio losses

$$L_n = \sum_{i \in \Lambda_n} U(i). \tag{3.8}$$

We make the following assumptions. Conditional on the liquidity state $r \in \{0, 1\}$ of a firm, losses are independent. The conditional distribution $M_r$ of losses with respect to a firm in liquidity state $r$ depends only on $r$. We suppose that losses are supported in a bounded interval on $\mathbb{R}_+$. We take $M_r$ as given and let $l_r = \int w \; M_r(dw)$ denote the expected loss caused by a firm in liquidity state $r$. For high-liquidity firms the probability of (large) losses is small relative to firms in the low-liquidity state. $M_1$ is more concentrated on large values than $M_0$. Specifically, we might assume that $M_1$ stochastically dominates $M_0$, i.e. for all bounded increasing functions $f : \mathbb{R}_+ \to \mathbb{R}$: $\int f dM_1 \geq \int f dM_0$. We however only suppose that $l_1 > l_0$.

Throughout, we will focus on the situation where the business partner network is dense ($d > 2$).

### 3.4.1 Deterministic Conditional Losses

We begin our analysis of aggregate portfolio losses $L_n$ in this section under the simplifying assumption that credit losses $U(i)$ depend deterministically on the liquidity state of firm $i$. Specifically, we simply set $M_r = \delta_r$ for $r \in \{0, 1\}$. This implies that the institution suffers no loss from positions with high-liquidity firms (where $r = 0$), and a loss of one unit of account from positions with low-liquidity firms ($r = 1$).

Let $\mu = \int_0^1 \nu_\rho Q(d\rho)$ be an equilibrium distribution of liquidity states. Here, the measures $\nu_\rho$ ($\rho \in [0, 1]$) are the extremal invariant measures of the voter model, and

$Q$ is the distribution of the random empirical proportion of low-liquidity firms in the economy which we denote by $\bar{\rho}$. Consider now the average loss $|\Lambda_n|^{-1}L_n$ in portfolio $\Lambda_n$. Since the measure $\nu_\rho$ $(\rho \in [0,1])$ are ergodic, we obtain by a conditional law of large numbers,

$$\lim_{n\to\infty} \frac{L_n}{|\Lambda_n|} = \bar{\rho} \tag{3.9}$$

$\mu$-almost surely. Even with deterministic conditional loss amounts not all loss uncertainty averages out. There is still uncertainty concerning average portfolio losses governed by the distribution $Q$. This distribution captures the systematic risk in the economy.

The average portfolio loss is thus not governed by the interaction of the firms, but simply by the distribution $Q$. This is due to the ergodicity of the extremal invariant measures of the voter model. The result (3.9) relies only on the validity of a law of large numbers and *not* on the specific structure of the ergodic measures in the decomposition of $\mu$. Whenever the ergodic measures have the correct one-dimensional marginal distribution, equation (3.9) holds.

Let us illustrate this in the benchmark case of conditionally independent firms. If we replace $\nu_\rho$ by a product measure $\pi_\rho$ of Bernoulli distributions with parameter $\rho$ and consider a distribution $\hat{\mu} = \int_0^1 \pi_\rho Q(d\rho)$ of liquidity states, contagion is not present any more. The mixture $\hat{\mu}$ corresponds to an economy in which the liquidity states of individual firms are not interdependent via direct business relations – they are only coupled through systematic risk captured by the distribution $Q$. In this case, equation (3.9) is still valid $\hat{\mu}$-almost surely.

Contagion does not affect the average per capita loss in the economy, but it increases the risk of large losses in *finite* portfolios. This effect can be quantified by a non-classical limit theorem which we state below.

We start with the special case where $Q = \delta_\rho$ for fixed $\rho \in (0,1)$ and investigate the portfolio losses associated with the extremal invariant distribution $\nu_\rho$ of the liquidity states. The case of general $Q$ is considered later.

***Theorem 3.4.1.*** *Let $d > 2$ and $Q = \delta_\rho$ for $\rho \in (0,1)$. Suppose additionally that $M_r = \delta_r$ for $r \in \{0,1\}$. For large portfolios the law of the losses $L_n$ can be approximated by a normal distribution $\mathcal{N}$:*

$$|\Lambda_n|^{-\frac{d+2}{2d}} \cdot (L_n - |\Lambda_n| \cdot \rho) = |\Lambda_n|^{-\frac{d+2}{2d}} \cdot \sum_{i\in\Lambda_n} (\xi(i) - \rho) \xrightarrow{w} \mathcal{N}(0,\sigma^2), \tag{3.10}$$

*where the limiting variance $\sigma^2 = \sigma^2(d)$ is given by*

$$\sigma^2 = \rho(1-\rho) \cdot \frac{\gamma_d \cdot d}{2^{d+3}\pi^{d/2}} \cdot \Gamma\left(\frac{d-2}{2}\right) \cdot \int_{[-1,1]^d} \int_{[-1,1]^d} \frac{1}{\|x-y\|_2^{d-2}} dxdy. \tag{3.11}$$

*Here $\Gamma$ is the Gamma-function and $\gamma = \gamma_d$ is given by*

$$\frac{1}{\gamma} = (2\pi)^{-d} \int_{(-\pi,\pi)^d} \left(1 - \frac{1}{d}\sum_{m=1}^{d} \cos x_m\right)^{-1} dx. \tag{3.12}$$

*The loss distribution can uniformly be approximated:*

$$\sup_{x\in\mathbb{R}_+} \left| \nu_\rho(L_n \geq x) - \Phi\left(\frac{|\Lambda_n|^{1/2}\rho - |\Lambda_n|^{-1/2}x}{\sigma \cdot |\Lambda_n|^{1/d}}\right)\right| \leq \epsilon_n, \tag{3.13}$$

*where $\epsilon_n \to 0$ as $n \to \infty$, and $\Phi$ is the standard normal distribution function.*

*Proof.* From Theorem 1 in Zähle (2001) we can derive the following normal approximation result for the basic voter model in dimension $d > 2$: Let $p(i)$ be the transition probability of the first step of a simple random walk starting at 0, i.e. $p(i) = (2d)^{-1}$ if $i = \pm e_j$ where $e_j$ is the $j$th unit vector, $j = 1, 2, \ldots, d$. Let $Z = (Z_1, \ldots, Z_d)$ be a random vector distributed according to $p$. The second moments of $Z$ are given by

$$Q_{l,k} = E[Z_l Z_k] = \frac{1}{d}\,\delta_{l,k},$$

where $\delta_{l,k}$ denotes the Kronecker symbol. Let $Q = (Q_{l,k})$ be the matrix of the second moments, and $|Q|$ the determinate of $Q$. Since $Q$ is invertible, we can define the quadratic form

$$\bar{Q}(x) = x^T Q^{-1} x.$$

Denoting the identity matrix in $\mathbb{R}^{d\times d}$ by $I$, we get $Q = \frac{1}{d}I$, $Q^{-1} = d \cdot I$, and $|Q| = d^{-d}$.

From a general result of Zähle (2001) for the linear voter model it follows in the particular case of the basic voter model for any Schwartz function $\phi : \mathbb{R}^d \to \mathbb{R}$ that as $r \to \infty$

$$r^{-(d+2)/2} \sum_{i\in\mathbb{Z}^d} [\xi(i) - \rho] \cdot \phi\left(\frac{i}{r}\right) \xrightarrow{w} \mathcal{N}(0, C_\rho B(\phi,\phi)). \tag{3.14}$$

Here $B$ is the bilinear functional on the Schwartz space $\mathcal{S}$ given by

$$B(\phi,\psi) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{\phi(x)\psi(y)}{\bar{Q}(x-y)^{(d-2)/2}} dxdy.$$

The multiplicative constant $C_\rho$ is defined by

$$C_\rho = \rho(1-\rho)\frac{\gamma}{2\pi^{d/2}|Q|^{1/2}}\Gamma\left(\frac{d-2}{2}\right),$$

where $\Gamma$ denotes the Gamma function and $\gamma$ is the escape probability of a discrete time simple random walk in dimension $d$ which starts in 0. It can be shown that the result holds also if $\phi$ is chosen to be the indicator of a box, e.g.

$$\phi = 1_{[-1,1]^d}.$$

Hence as $n \to \infty$ we get weak convergence,

$$\begin{aligned} |\Lambda_n|^{-\frac{d+2}{2d}} \sum_{i \in \Lambda_n} (\xi(i) - \rho) &= (2n+1)^{-\frac{d+2}{2}} \sum_{i \in \mathbb{Z}^d} (\xi(i) - \rho) 1_{[-1,1]^d} \left(\frac{i}{n}\right) \\ \xrightarrow{w} \quad \mathcal{N}\left(0, \frac{C_\rho B(1_{[-1,1]^d}, 1_{[-1,1]^d})}{2^{d+2}}\right). \end{aligned}$$

In order to verify the approximation result (3.10), we have to calculate the asymptotic variance. It follows that

$$\begin{aligned} B(1_{[-1,1]^d}, 1_{[-1,1]^d}) &= \frac{1}{d^{(d-2)/2}} \int_{[-1,1]^d} \int_{[-1,1]^d} \|x - y\|_2^{-(d-2)} dx dy, \\ C_\rho &= \rho(1-\rho) \frac{\gamma}{2\pi^{d/2} \cdot d^{-d/2}} \cdot \Gamma\left(\frac{d-2}{2}\right). \end{aligned}$$

It is now easy to see that

$$\sigma^2 = \frac{C_\rho B(1_{[-1,1]^d}, 1_{[-1,1]^d})}{2^{d+2}}, \tag{3.15}$$

which proves (3.11).

Next we derive the uniform approximation (3.13). Since the distribution function of the normal distribution is continuous, it follows from Exercise 2.6. in Chapter 2 of Durrett (1996) that

$$\sup_{z \in \mathbb{R}} \left| \nu_\rho \left( \frac{|\Lambda_n|^{-\frac{d+2}{2d}} (L_n - |\Lambda_n|\rho)}{\sigma} \geq z \right) - \Phi(-z) \right| \leq \epsilon_n,$$

where $\epsilon_n \to 0$ as $n \to \infty$. Approximation (3.13) follows immediately, since

$$\{L_n \geq x\} = \left\{ \frac{|\Lambda_n|^{-\frac{d+2}{2d}} (L_n - |\Lambda_n|\rho)}{\sigma} \geq \frac{|\Lambda_n|^{-\frac{d+2}{2d}} (x - |\Lambda_n|\rho)}{\sigma} \right\}.$$

Finally, rewrite the last term in the last bracket:

$$\frac{|\Lambda_n|^{-\frac{d+2}{2d}} (x - |\Lambda_n|\rho)}{\sigma} = \frac{|\Lambda_n|^{-1/2} x - |\Lambda_n|^{1/2} \rho}{\sigma \cdot |\Lambda_n|^{1/d}}.$$

The formula for the escape probability $\gamma = \gamma_d$ is derived in Theorem 3.4.2. $\square$

The escape probability $\gamma_d$ can be calculated as follows.

***Theorem 3.4.2.*** *Let $Y_n$ be a simple random walk on $\mathbb{Z}^d$ with $d \geq 3$. The escape probability $\gamma_d$ can be calculated by*

$$\gamma_d = \frac{1}{J(d)}, \tag{3.16}$$

*where the quantity $J(d)$ is defined by*

$$J(d) = (2\pi)^{-d} \int_{(-\pi,\pi)^d} \left(1 - \frac{1}{d}\sum_{m=1}^{d} \cos x_m\right)^{-1} dx. \tag{3.17}$$

*Numerical values are given in the following table:*

| $d$ | $J(d)$ | $\gamma_d$ |
|---|---|---|
| 3 | 1.516386 | 0.659463 |
| 4 | 1.239467 | 0.806798 |
| 5 | 1.156308 | 0.864821 |
| 6 | 1.116963 | 0.895285 |

*Proof.* The proof uses arguments from Chapter 3 of Durrett (1996). We denote by $\tau_n$ the optional time of the $n$th return of $Y_n$ to 0. Note that $P(\tau_1 < \infty) < 1$ by the transience of the random walk. Hence, we have

$$\begin{aligned} &\sum_{m=0}^{\infty} P(Y_m = 0) = \sum_{n=0}^{\infty} P(\tau_n < \infty) \\ = &\sum_{n=0}^{\infty} P(\tau_1 < \infty)^n = \frac{1}{1 - P(\tau_1 < \infty)} = \frac{1}{\gamma_d}. \end{aligned}$$

Thus,

$$\gamma_d = \left(\sum_{m=0}^{\infty} P(Y_n = 0)\right)^{-1}$$

It remains to be shown that

$$\sum_{m=0}^{\infty} P(Y_n = 0) = J(d).$$

By $\phi$ we denote the characteristic function of one step of the random walk, i.e.

$$\phi(x) = E(\exp(ixY_1)) = \frac{1}{d}\sum_{j=1}^{d} \cos x_j$$

where the last identity follows from Euler's formula. The independence of the increments of the random walk implies that

$$\phi^n(x) = E(\exp(ixY_n))$$

Since $Y_n$ is $\mathbb{Z}^d$-valued, we have the following simple Fourier inversion:

$$P(Y_n = 0) = (2\pi)^{-d} \int_{(-\pi,\pi)^d} \phi^n(x) dx$$

For $r \in (0, 1)$, it holds that

$$\left| \sum_{n=0}^{\infty} r^n \phi^n(x) \right| \leq \sum_{n=0}^{\infty} r^n \|\phi\|_\infty^n = \sum_{n=0}^{\infty} r^n = \frac{1}{1-r}$$

By the bounded convergence theorem, we get

$$\sum_{n=0}^{\infty} r^n P(Y_n = 0) = (2\pi)^{-d} \int_{(-\pi,\pi)^d} \frac{1}{1 - r\phi(x)} dx$$

Observe that $\phi$ is a real function. The term

$$\frac{1}{1 - r\phi(x)}$$

is bounded between 0 and 1 if $\phi(x) \leq 0$, and increases to $(1 - \phi(x))^{-1}$ as $r \nearrow 1$ if $\phi(x) > 0$. Hence, the monotone and bounded convergence theorem imply that

$$\sum_{n=0}^{\infty} P(Y_n = 0) = (2\pi)^{-d} \int_{(-\pi,\pi)^d} \left( 1 - \frac{1}{d} \sum_{m=1}^{d} \cos x_m \right)^{-1} dx$$

In order to obtain the numerical values in the table observe that

$$J(d) = d \cdot I(d; 1) = L(d; 1) + 1$$

where the functions $I$ and $L$ are defined and evaluated in Kondo and Hara (1987). $\square$

The re-scaling in (3.10) is non-classical. This is caused by the strong dependence in the equilibrium distribution of liquidity states, which results from the contagion dynamics. Unfortunately, we are not able to provide bounds of Berry-Esseen-type for the errors $\epsilon_n$ in (3.13), which would help to understand the speed of convergence.

By inequality (3.13) the probability of a loss larger than $x \in \mathbb{R}_+$ can uniformly be approximated by the function

$$\Psi_{d,\rho}(|\Lambda_n|, x) = \Phi \left( \frac{|\Lambda_n|^{1/2}\rho - |\Lambda_n|^{-1/2} x}{\sigma(d) \cdot |\Lambda_n|^{1/d}} \right), \tag{3.18}$$

where $|\Lambda_n| = (2n+1)^d$ is the size of the portfolio $\Lambda_n = [-n, n]^d$. Heuristically, interpolation between sizes of the portfolios $\Lambda_n$ allows us to define the approximate loss probabilities larger than $x \in \mathbb{R}_+$ for portfolio size $u \in \mathbb{R}_+$ by

$$\Psi_{d,\rho}(u, x) = \Phi\left(\frac{u^{1/2}\rho - u^{-1/2}x}{\sigma(d) \cdot u^{1/d}}\right). \tag{3.19}$$

Hence, losses of a portfolio of $u$ firms are approximately normal with mean $\rho u$ and variance $\sigma^2(d)u^{1+\frac{2}{d}}$, that is, the losses of $u$ firms are approximately $\mathcal{N}(\rho u, \sigma^2(d)u^{1+\frac{2}{d}})$. The risk of large losses is captured by the variance of the approximating normal variable. The variance is of order $u^{1+\frac{2}{d}}$. The exponent decreases to 1 as $d \to \infty$.

The interaction of the firms leads to strong dependence of the liquidity states of different firms. We shall compare the results for the distribution $\nu_\rho$ to the benchmark case of independent firms. That is, we will assume that the benchmark distribution $\pi_\rho$ of liquidity states is given by a product of Bernoulli measures with parameter $\rho$. If we exchange $\nu_\rho$ against $\pi_\rho$, we have to replace the normalization factor $|\Lambda_n|^{-\frac{d+2}{2d}}$ in (3.10) simply by the usual $|\Lambda_n|^{-\frac{1}{2}}$ and use instead of the limiting variance $\sigma^2(d)$ the quantity $\rho(1-\rho)$. The uniform approximation (3.13) becomes in this case

$$\sup_{x \in \mathbb{R}_+} \left| \pi_\rho(L_n \geq x) - \Phi\left(\frac{|\Lambda_n|^{1/2}\rho - |\Lambda_n|^{-1/2}x}{\sqrt{\rho(1-\rho)}}\right) \right| \leq \epsilon_n, \tag{3.20}$$

where $\epsilon_n \to 0$ as $n \to \infty$. For independent firms the speed of convergence to the normal distribution can be bounded by the Berry-Esseen theorem (see e.g. Theorem 4.9. and Remarks in Chapter 2 of Durrett (1996)):

$$\epsilon_n \leq \frac{1 + 2\rho(\rho - 1)}{\sqrt{\rho(1-\rho)}} \cdot \frac{1}{(2n)^{d/2}}. \tag{3.21}$$

By inequality (3.20) the probability of a loss larger than $x \in \mathbb{R}_+$ can uniformly be approximated by the function

$$\hat{\Psi}_\rho(|\Lambda_n|, x) = \Phi\left(\frac{|\Lambda_n|^{1/2}\rho - |\Lambda_n|^{-1/2}x}{\sqrt{\rho(1-\rho)}}\right). \tag{3.22}$$

Again interpolation between sizes of the portfolios $\Lambda_n$ allows us to define the approximate loss probabilities larger than $x \in \mathbb{R}_+$ for portfolio size $u \in \mathbb{R}_+$ by

$$\hat{\Psi}_\rho(u, x) = \Phi\left(\frac{u^{1/2}\rho - u^{-1/2}x}{\sqrt{\rho(1-\rho)}}\right). \tag{3.23}$$

Hence, losses of a portfolio of $u$ firms are approximately normal with mean $\rho u$ and variance $\rho(1-\rho)u$, that is, the losses of $u$ firms are approximately $\mathcal{N}(\rho u, \rho(1-\rho)u)$. In contrast to the contagion case, the order of the variance is simply $u$.

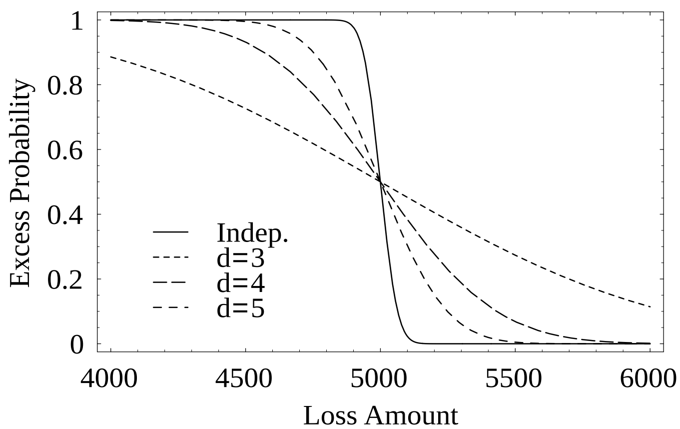


*Figure 3.1: Probability of a portfolio loss exceeding a given amount, varying the degree $d$ of connectedness ($u = 10.000$ and $\rho = 0.5$).*

The order of the variance relates to the the riskiness of a portfolio. With contagion, portfolios are more risky than without contagion. In the case of contagion, the order of the variance is $u^{1+\frac{2}{d}}$. The exponent decreases as $d$ increases. Thus, the portfolio becomes more risky if $d$ is small. For $d \leq 5$ and reasonable portfolio sizes, say 10.000 firms, this effect cannot be neglected.

To illustrate this, we consider a portfolio of size $u = 10.000$ with parameter $\rho = 0.5$, i.e. the marginal probability that a firm is in the low-liquidity state is 0.5. In Figures 3.1 and 3.2 we plot the approximate loss distribution for the benchmark case and the contagion case, where for the latter we vary the denseness $d$ of the business partner network. As expected, in comparison with the independence (benchmark) case the loss distribution exhibits a higher variance when credit contagion phenomena are present. Put another way, firm interaction leads to the portfolio being more risky in terms of large losses. With interaction, the probability of exceeding a given loss amount above average losses is larger than in the independence case.

The difference in loss probabilities depends on the denseness of the business partner network. The *lower* $d$, the more volatile is the loss distribution. In this sense, the economy with the least dense network (here $d = 3$) induces the riskiest portfolios. The higher $d$, the lower is the likelihood of large losses. The rationale of this was mentioned in Section 3.3.2: the denser the network, the lower is the degree of interaction and therefore the less likely is a large loss.

The non-trivial equilibrium distribution depends on the Bernoulli parameter $\rho$ of

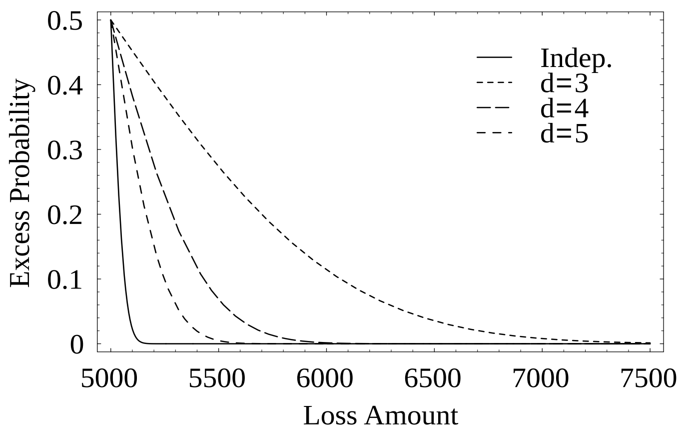


*Figure 3.2: Probability of a portfolio loss exceeding a given amount, varying the degree $d$ of connectedness ($u = 10.000$ and $\rho = 0.5$).*

the one-dimensional marginal distribution of the liquidity states – and so does the loss distribution. The dependence of loss probabilities on $\rho$ for a given $d$ is shown in Figure 3.3: the lower $\rho$, the lower is the likelihood of large losses.

The approximate loss density for benchmark and interaction case (in dependence of $d$) is shown in Figure 3.4. While in case $Q = \delta_\rho$ all loss uncertainty averages out in infinite portfolios (cf. (3.9)), for finite portfolios losses fluctuate around $u \cdot \rho = 5000$. Corresponding to our discussion above, with interdependent liquidity states losses fluctuate more excessively when compared with the benchmark case, where the degree of fluctuation decreases with increasing network denseness $d$.

Having investigated the loss distribution in the special case where $Q = \delta_\rho$ for $\rho \in (0, 1)$, we now consider the case of general $Q$. In this situation the invariant distributions $\mu$ of liquidity states are mixtures of the extremal measures $\nu_\rho$, which we focused on in the special case (for a given $\rho$). Let $\mu = \int_0^1 \nu_\rho Q(d\rho)$ be an equilibrium liquidity distribution. If $Q$ puts positive mass on 0 or 1, all firms are in the same liquidity state with positive probability. In order to avoid unnecessary technical complications, we exclude these trivial cases as before and assume $Q(\{0\}) = Q(\{1\}) = 0$. In this general case, the exact probability of a loss larger than $x \in \mathbb{R}_+$ equals

$$\int \nu_\rho(L_n \geq x)Q(d\rho).$$

In a large portfolio, the law of the losses $L_n$ can be uniformly approximated by a mixture of normal distributions:

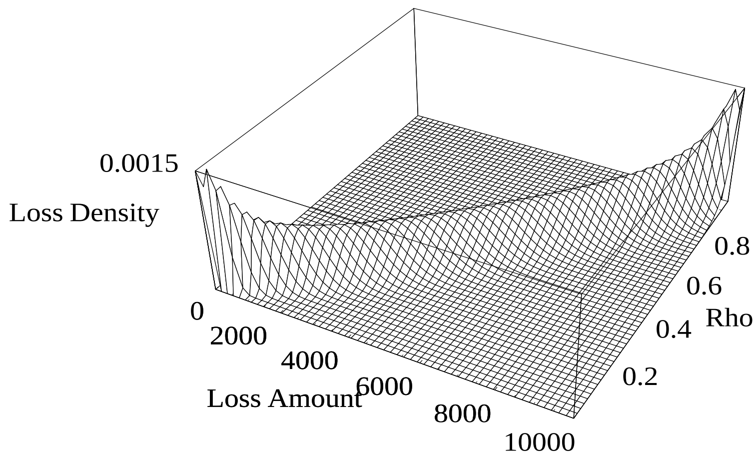


*Figure 3.3: Approximate density of portfolio losses, varying the marginal parameter $\rho$ ($u = 10.000$ and $d = 3$).*

***Corollary 3.4.3.*** *Let $d > 2$ and $M_r = \delta_r$ for $r \in \{0, 1\}$. The distribution of portfolio losses $L_n$ can uniformly be approximated, i.e.*

$$\sup_{x\in\mathbb{R}} \left| \int \nu_\rho(L_n \geq x) Q(d\rho) - \int \Phi\left( \frac{|\Lambda_n|^{1/2}\rho - |\Lambda_n|^{-1/2}x}{\sigma(\rho)|\Lambda_n|^{1/d}} \right) Q(d\rho) \right| \leq \epsilon_n, \tag{3.24}$$

*where the error bound $\epsilon_n \to 0$ as $n \to \infty$.*

*Proof.* The distribution of

$$|\Lambda_n|^{-\frac{d+2}{2d}} \sum_{i\in\Lambda_n} (\xi(i) - \rho)$$

under the measure $\nu_\rho$ will be denoted by $\varsigma_\rho^n$. We define the quantity

$$\delta_\rho^n := \sup_{n'\geq n} \sup_{z\in\mathbb{R}} \left| \varsigma_\rho^{n'}([z, \infty)) - \Phi\left( -\frac{z}{\sigma(\rho)} \right) \right|.$$

Inequality (3.13) implies that $\delta_\rho^n$ converges to 0 for all $\rho \in (0, 1)$ as $n \to \infty$. Observe that $\rho \mapsto \delta_\rho^n$ is measurable. For $\epsilon > 0$ we can therefore define measurable sets

$$A_\epsilon^n = \{\rho \in (0, 1) : \delta_\rho^n < \epsilon\}.$$

Then $A_\epsilon^n \subseteq A_\epsilon^{n+1}$, and $Q(A_\epsilon^n) \nearrow 1$ as $n \to \infty$. Choose $n_0$ large enough such that

$$Q(A_\epsilon^{n_0}) \geq 1 - \epsilon.$$

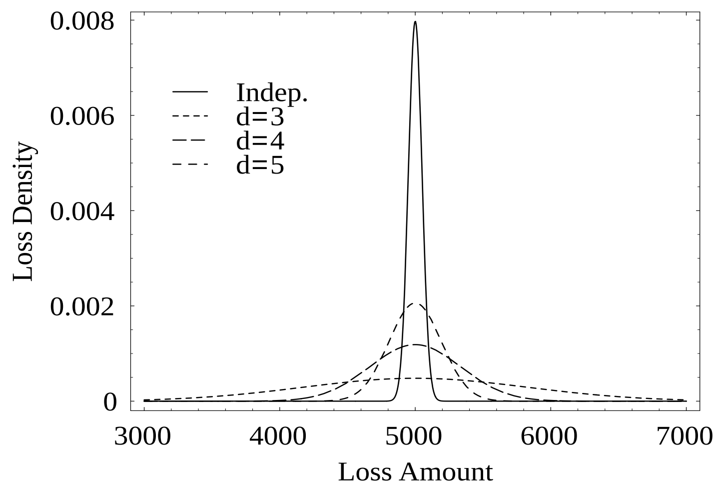


*Figure 3.4: Approximate density of portfolio losses, varying the degree $d$ of connectedness ($u = 10.000$ and $\rho = 0.5$).*

Let $\rho \mapsto z(\rho)$ be a measurable mapping. Then for all $n \geq n_0$ we get

$$\begin{aligned} & \left| \int \left[ \varsigma_\rho^n([z(\rho), \infty)) - \Phi\left(-\frac{z(\rho)}{\sigma(\rho)}\right)\right] Q(d\rho) \right| \\ \leq \; & 2(1 - Q(A_\epsilon^n)) + \sup_{\rho \in A_\epsilon^n} \sup_{z' \in \mathbb{R}} \left| \varsigma_\rho^n([z', \infty)) - \Phi\left(-\frac{z'}{\sigma(\rho)}\right)\right| \\ \leq \; & 3\epsilon \end{aligned}$$

Let $x \in \mathbb{R}$ be arbitrary, and let $n \geq n_0$. We can choose

$$z(\rho) = |\Lambda_n|^{-\frac{d+2}{2d}}(x - |\Lambda_n|\rho).$$

It follows that for any $x \in \mathbb{R}$ and $n \geq n_0$ the following inequality holds

$$\left| \int \nu_\rho(L_n \geq x) Q(d\rho) - \int \Phi\left(\frac{|\Lambda_n|^{1/2}\rho - |\Lambda_n|^{-1/2}x}{\sigma(\rho)|\Lambda_n|^{1/d}}\right) Q(d\rho) \right| \leq 3\epsilon.$$

□

Based on this result, in close analogy to (3.19) interpolation between sizes of the portfolios $\Lambda_n$ allows us to define the approximate loss probabilities larger than $x \in \mathbb{R}_+$ for portfolio size $u \in \mathbb{R}_+$ by

$$\int \Phi\left(\frac{u^{1/2}\rho - u^{-1/2}x}{\sigma(\rho) \cdot u^{1/d}}\right) Q(d\rho).$$

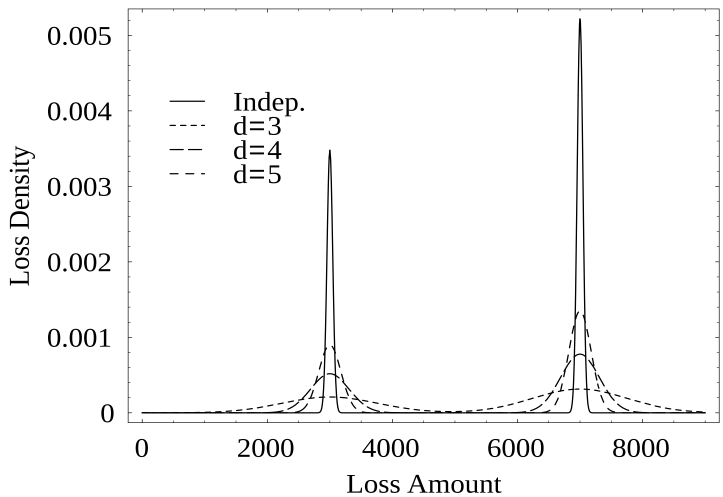


*Figure 3.5: Approximate density of portfolio losses, varying the degree $d$ of connectedness ($u = 10.000$ and $Q = 0.4\delta_{0.3} + 0.6\delta_{0.7}$).*

Paralleling (3.23), in the benchmark case with independent firms the approximate loss probabilities can be defined by

$$\int \Phi\left(\frac{u^{1/2}\rho - u^{-1/2}x}{\sqrt{\rho(1-\rho)}}\right) Q(d\rho), \quad x, u \in \mathbb{R}_+.$$

In both cases – with and without contagion – the systematic risk described by the distribution $Q$ governs the approximate loss distribution. The Gaussian integrands cause additional fluctuations around their random means. If contagion is present, the variance of these Gaussian distributions is of larger order in the number of positions $u$. The order decreases if the business partner network becomes denser.

In Figure 3.5 we illustrate the approximate density of portfolio losses in the case $Q = 0.4\delta_{0.3} + 0.6\delta_{0.7}$. The portfolio size is again $u = 10.000$. In infinite portfolios, according to the distribution $Q$ average losses $\rho = 0.3$ with probability 0.4 and $\rho = 0.7$ with probability 0.6. In finite portfolios losses fluctuate around $0.3 \cdot u = 3000$ (with probability 0.4) and $0.7 \cdot u = 7000$ (with probability 0.6), as prescribed by $Q$. In analogy to the no-uncertainty case $Q = \delta_\rho$ considered in Figure 3.4, interaction leads to more fluctuations when compared to the benchmark case. The degree of additional fluctuation depends on the denseness of the network.

### 3.4.2 Stochastic Conditional Losses

In this section we study the distribution of aggregate portfolio losses $L_n$ in the general case, i.e. without a particular assumption on the shape of the conditional distribution $M_r$.

In a first step we consider the average losses in the portfolio $\Lambda_n$. Let $\mu = \int_0^1 \nu_\rho Q(d\rho)$ again be an equilibrium distribution of liquidity states where the average number of low-liquidity firms in the whole economy is distributed according to $Q$. The joint distribution of losses is given by the mixture

$$\beta(dw) = \int \left(\otimes_{i\in\mathbb{Z}^d} M_{\xi(i)}\right)(dw)\mu(d\xi), \quad w \in \mathbb{R}^{\mathbb{Z}^d}.$$

As with deterministic conditional losses, by a conditional law of large numbers we have that

$$\lim_{n\to\infty} \frac{L_n}{|\Lambda_n|} =: m \tag{3.25}$$

exists $\beta$-almost surely. Writing $m = \rho(l_1 - l_0) + l_0$, we obtain that $\rho$ is random with distribution $Q$. Due to the ergodicity of the measures $\nu_\rho$, in infinite portfolios average losses do not depend on the interaction of firms, but only on systematic risk. Our next result shows that in large portfolios the quantiles $q_\alpha(L_n)$ of the loss distribution are essentially governed by the quantiles of $Q$.

***Proposition 3.4.4.*** *Let $q_\alpha(Q)$ be the $\alpha$-quantile of the distribution $Q$ and assume that the cumulative distribution function $G$ of $Q$ is strictly increasing at $q_\alpha(Q)$, i.e. $G(q_\alpha(Q)+\epsilon) > \alpha$ and $G(q_\alpha(Q)-\epsilon) < \alpha$ for every $\epsilon > 0$. Then*

$$\lim_{n\to\infty} \frac{q_\alpha(L_n)}{|\Lambda_n|} = q_\alpha(Q)(l_1 - l_0) + l_0.$$

*where $l_r$ is the expected loss on a position with a firm in liquidity state $r \in \{0,1\}$. Here, $q_\alpha(L_n)$ denotes an $\alpha$-quantile of the distribution of $L_n$ under the measure $\beta$.*

*Proof.* For $\rho \in [0,1]$ define the probability measures

$$\beta_\rho(dw) = \int \left(\otimes_{i\in\mathbb{Z}^d} M_{\xi(i)}\right)(dw)\nu_\rho(d\xi), \quad w \in \mathbb{R}^{\mathbb{Z}^d}.$$

First observe that, due to (3.25),

$$\lim_{n\to\infty} \beta_\rho\left(\frac{\frac{L_n}{|\Lambda_n|} - l_0}{l_1 - l_0} \le a\right) = \begin{cases} 1 & , \quad \rho < a \\ 0 & , \quad \rho > a \end{cases}$$

Let $\epsilon > 0$ and let $G$ be the cumulative distribution function of $Q$. Then

$$\begin{aligned}
\limsup_{n\to\infty}\ & \beta\big\{L_n - l_0|\Lambda_n| \le |\Lambda_n|(l_1-l_0)(q_\alpha(Q)-\epsilon)\big\} \\
&= \limsup_{n\to\infty} \int_0^1 \beta_\rho\big\{L_n - l_0|\Lambda_n| \le |\Lambda_n|(l_1-l_0)(q_\alpha(Q)-\epsilon)\big\} dG(\rho) \\
&\le \int_0^1 \limsup_{n\to\infty} \beta_\rho\left(\frac{\frac{L_n}{|\Lambda_n|}-l_0}{l_1-l_0} \le q_\alpha(Q)-\epsilon\right) dG(\rho) \\
&\le \int_0^1 1_{(-\infty,q_\alpha(Q)-\epsilon]}(\rho) dG(\rho) \\
&= G(q_\alpha(Q)-\epsilon)\ < \alpha,
\end{aligned}$$

where the last equality is strict by assumption. The first inequality follows from Fatou's lemma. Analogously,

$$\liminf_{n\to\infty}\ \beta\big\{L_n - l_0|\Lambda_n| \le |\Lambda_n|(l_1-l_0)(q_\alpha(Q)+\epsilon)\big\} \ge G(q_\alpha(Q)+\epsilon/2) > \alpha.$$

Hence, for $n$ large enough:

$$|\Lambda_n|(l_1-l_0)(q_\alpha(Q)-\epsilon) \le q_\alpha(L_n - l_0|\Lambda_n|) \le |\Lambda_n|(l_1-l_0)(q_\alpha(Q)+\epsilon).$$

The claim follows from observing that $q_\alpha(L_n - l_0|\Lambda_n|) = q_\alpha(L_n) - l_0|\Lambda_n|$. □

Frey and McNeil (2001) proved a similar result for exchangeable Bernoulli mixture models, in which credit losses are conditionally independent given some exogenous macro-economic factors. In this context the quantiles of the given factor distribution (the mixing distribution) essentially determine the quantiles of the loss distribution for large homogeneous portfolios. This tail behavior is of central significance for risk measurement and management, as it corresponds to a probabilistic assessment of the scenarios with extremely large losses. Analogously, in our credit contagion approach the tail properties of the systematic risk $Q$ essentially govern the tail behavior of aggregate losses in large portfolios, i.e. the extent of excessive fluctuations of the losses $L_\infty$ in an infinitely large portfolio.

We again focus first on the case $Q = \delta_\rho$ for $\rho \in (0,1)$, and investigate the distribution of the losses

$$\beta(dw) = \int \left(\otimes_{i\in\mathbb{Z}^d} M_{\xi(i)}\right)(dw)\nu_\rho(d\xi), \quad w \in \mathbb{R}^{\mathbb{Z}^d}.$$

Like in the case of deterministic conditional losses, for large portfolios the law of the losses $L_n$ can again be approximated by a normal distribution. In this case, the expected loss equals $m$ as defined in (3.25) and can be written as

$$m = \rho(l_1 - l_0) + l_0.$$

From Theorem 3.4.1 we can derive the weak convergence of the losses in the stochastic case:

***Theorem 3.4.5.*** *Let $d > 2$ and suppose that $Q = \delta_\rho$ for $\rho \in (0,1)$. For large portfolios the distribution of losses satisfies*

$$|\Lambda_n|^{-\frac{d+2}{2d}} \cdot (L_n - |\Lambda_n| \cdot m) = |\Lambda_n|^{-\frac{d+2}{2d}} \cdot \sum_{i \in \Lambda_n} (U(i) - m) \xrightarrow{w} \mathcal{N}(0, (l_1 - l_0)^2 \cdot \sigma^2)$$

*where the $\sigma^2$ denotes the limiting variance (3.11). The loss distribution can uniformly be approximated, i.e.*

$$\sup_{x \in \mathbb{R}} \left| \beta(L_n \geq x) - \Phi\left( \frac{|\Lambda_n|^{1/2} m - |\Lambda_n|^{-1/2} x}{(l_1 - l_0)\sigma \cdot |\Lambda_n|^{1/d}} \right) \right| \leq \epsilon_n, \tag{3.26}$$

*where $\epsilon_n \to 0$ as $n \to \infty$.*

*Proof.* This is a corollary of the normal approximation results in the deterministic case. Define the function $f : \{0,1\} \to \{l_0, l_1\}$ by $f(0) = l_0$ and $f(1) = l_1$. $f$ is used to introduce the random variables $m_i = f(\xi(i))$, $i \in \mathbb{Z}^d$. It is easy to see that (3.10) implies

$$|\Lambda_n|^{-\frac{d+2}{2d}} \sum_{i \in \Lambda_n} (m_i - m) \xrightarrow{w} \mathcal{N}(0, (l_1 - l_0)^2 \cdot \sigma^2). \tag{3.27}$$

Denote now by $(X_{r,i})_{i \in \mathbb{Z}^d}$ independent random variables with distribution $M_r$, $r \in \{0,1\}$. Then we can rewrite the renormalized losses as

$$\begin{aligned} |\Lambda_n|^{-\frac{d+2}{2d}} (L_n - |\Lambda_n| m) &= |\Lambda_n|^{-\frac{d+2}{2d}} \sum_{i \in \Lambda_n,\ \xi(i)=0} (X_{0,i} - m_0) \\ &+ |\Lambda_n|^{-\frac{d+2}{2d}} \sum_{i \in \Lambda_n,\ \xi(i)=1} (X_{1,i} - m_1) \\ &+ |\Lambda_n|^{-\frac{d+2}{2d}} \sum_{i \in \Lambda_n} (m_i - m) \end{aligned}$$

The last summand on the right hand side converges weakly according to (3.27). The other two terms converge almost surely to 0; w.l.o.g. we will prove this fact only for the first term, i.e.

$$|\Lambda_n|^{-\frac{d+2}{2d}} \sum_{i \in \Lambda_n,\ \xi(i)=0} (X_{0,i} - m_0) = |\Lambda_n|^{-\frac{d+2}{2d}} \sum_{i \in \Lambda_n,\ \xi(i)=0} (X_{0,i} - l_0). \tag{3.28}$$

The random number of summands in (3.28) equals $c(n) = |\{i \in \Lambda_n : \xi(i) = 0\}|$ and is almost surely increasing to $\infty$ as $n \to \infty$. Theorem 8.7. of Chapter 1 in Durrett (1996) implies for $\epsilon > 0$ that

$$c(n)^{-1/2}(\log c(n))^{-(1/2+\epsilon)} \sum_{i\in\Lambda_n,\ \xi(i)=0} (X_{0,i} - l_0) \tag{3.29}$$

converges to 0 as $n \to \infty$. The last result can also be viewed as a consequence of the law of iterated logarithm.

Now observe that for $\epsilon > 0$ the sequence $c(n)$ satisfies

$$\frac{c(n)^{1/2}(\log c(n))^{1/2+\epsilon}}{|\Lambda_n|^{\frac{d+2}{2d}}} \leq \frac{|\Lambda_n|^{1/2}(\log|\Lambda_n|)^{1/2+\epsilon}}{|\Lambda_n|^{\frac{d+2}{2d}}} = \frac{(\log|\Lambda_n|)^{1/2+\epsilon}}{|\Lambda_n|^{1/d}}.$$

The last term converges to 0 as $n \to \infty$. This fact together with (3.29) implies that the terms in (3.28) converge to 0 as $n \to \infty$.

Altogether we obtain for $n \to \infty$ the weak convergence,

$$|\Lambda_n|^{-\frac{d+2}{2d}} \cdot (L_n - |\Lambda_n| \cdot m) \xrightarrow{w} \mathcal{N}(0, (l_1 - l_0)^2 \cdot \sigma^2).$$

The uniform approximation (3.26) is obtained with the same arguments as in the deterministic case. $\square$

Based on inequality (3.26), interpolation between sizes of the portfolios $\Lambda_n$ allows us to define the approximate loss probabilities larger than $x \in \mathbb{R}_+$ for portfolio size $u \in \mathbb{R}_+$ by

$$\Phi\left(\frac{u^{1/2}m - u^{-1/2}x}{(l_1 - l_0)\sigma \cdot u^{1/d}}\right).$$

This result corresponds to formula (3.19) which we obtained in the case with deterministic conditional losses. In case of stochastic conditional losses the limiting variance is multiplied by the factor $(l_1 - l_0)^2$, which depends only on the expected value of the loss distributions $M_r$, $r \in \{0, 1\}$. Because of the non-classical re-scaling the random fluctuations of the distributions $M_r$ are averaged out in the normal approximation.

In analogy to Corollary 3.4.3, we extend our analysis of the loss distribution to general invariant distributions $\mu$ of liquidity states, which are mixtures of the extremal measures we have considered so far. The joint distribution of the losses is given by the mixture

$$\beta(dw) = \int \left(\otimes_{i\in\mathbb{Z}^d} M_{\xi(i)}\right)(dw)\mu(d\xi), \quad w \in \mathbb{R}^{\mathbb{Z}^d}.$$

***Corollary 3.4.6.*** *Let $d > 2$. For a large portfolio, the distribution of losses $L_n$ can uniformly be approximated:*

$$\sup_{x\in\mathbb{R}} \left| \beta(L_n \geq x) - \int \Phi\left( \frac{|\Lambda_n|^{1/2} m - |\Lambda_n|^{-1/2} x}{(l_1 - l_0)\sigma(\rho)|\Lambda_n|^{1/d}} \right) Q(d\rho) \right| \leq \epsilon_n, \tag{3.30}$$

*where $m = \rho(l_1 - l_0) + l_0$ and $\epsilon_n \to 0$ as $n \to \infty$.*

*Proof.* Analogous to the proof of Corollary 3.4.3. □

Compared with inequality (3.24), if conditional losses are stochastic the approximate variance $\sigma^2(\rho)$ is multiplied by a factor $(l_1 - l_0)^2$ and the averages of low liquidity states $\rho$ are replaced by $m$. Qualitatively, the approximate loss distributions has similar properties in both cases (3.24) and (3.30). Interestingly, the fluctuations of the distributions $M_r$ around their means are averaged out in the normal approximation; only the expectations enter the inequality (3.30).

## 3.5 Discussion

Credit contagion refers to the propagation of economic distress from one firm to another. A thorough understanding of contagion processes is essential for the management and regulation of financial institutions. In this chapter we model credit contagion phenomena and study their effects on the volatility of losses on large portfolios of financial positions. We provide an explicit approximation of aggregate losses.

Our contagion model is tractable, but stylized. The economy is modeled by a multi-dimensional lattice, whose nodes are identified with firms. The edges represent business partner relationships. The dimension of the lattice measures the denseness of the business partner network. Firms are homogeneous in their individual characteristics. While they may be in different states, they have the same number of business partner relationships and are of equal size.

The business partner relationships are the channel for credit contagion phenomena, i.e. the propagation of liquidity shocks through a network of obligations. The direction in which shocks are propagated is symmetric. The likelihood of a healthy firm to become distressed increases with the number of distressed business partners. Vice versa, the likelihood of a distressed firm to make a turnaround increases with the number of healthy business partners. Less realistic is, first, the special choice of the transition rates and, second, the symmetry in the business relationships: any two firms influence each other to the same degree. For future research, it would be very interesting to account for such asymmetries, e.g. by considering directed graphs, and to relax some of the assumptions in alternative case studies.

# Chapter 4

# A Continuous Time Approximation of an Evolutionary Stock Market Model

## 4.1 Introduction

The axiom of profit maximization is a cornerstone of neoclassical economics. Often it is justified by the market selection hypothesis, which argues that maximization describes the long-run market behavior induced by an evolutionary selection process, cf. Friedman (1953) and Fama (1965). While intuitively appealing, this argument clearly needs a rigorous analysis.

An explicit model for the market selection mechanism has been proposed in a seminal paper by Blume and Easley (1992). In an asset model with endogenous prices in discrete time, agents follow simple trading strategies. They keep the proportion of wealth invested in each asset fixed over time and reinvest their payoffs. The market process induces a redistribution of wealth among traders. Blume and Easley (1992) investigate the long-run dynamics of the selection process. Under strong conditions on the underlying random variables and the payoff structure of the assets they identify the unique survivor of the market selection process.

This result has recently been generalized. Evstigneev et al. (2002) extend the model to a more complex payoff structure for the case that uncertainty is modeled by a sequence of independent random variables taking values in a finite state space. They identify the unique surviving strategy. For general ergodic states, Hens and Schenk-Hoppé (2004) derive local stability results.

In the current chapter, we provide a continuous time approximation for the model of Blume and Easley (1992) for general random payoffs of the assets. Here, we assume that trading takes place at a higher frequency and that in each trading interval agents reinvest only a fraction of their wealth. If the payoffs of the assets converge nicely (as the time between two successive trading dates approaches zero), then also the wealth process of the agents converges by a functional limit theorem which is closely related to the well known Euler scheme. The continuous time limit of the wealth process equals the solution of a non linear integral equation in a random environment.

The continuous time approximation of the wealth process relies on a proper convergence of the payoffs of the assets, as the length of the trading intervals tends to zero. We suggest an economically meaningful model for the dividend processes and their convergence. Dividend payments are modeled as increments of stochastic firm value processes. Conditions on these processes are identified, which ensure the applicability of the functional limit theorem. For this purpose, the notion of locally finite kernels turns out to be useful.

In a further step, we analyze the long-run asymptotic behavior of the continuous time approximation in the simplest special case. Namely, we assume that the dividend process of the assets is deterministic and constant. The Markovian case will be investigated in Buchmann and Weber (2004b). For constant dividend payments, the deterministic dynamics of the wealth process in continuous time is described by a non linear, autonomous ordinary differential equation. We characterize its long-run asymptotic behavior. Here, we employ the technique of Lyapunov functions.

Finally, we compare these results to a Walrasian equilibrium of myopic agents who are price takers. In continuous time, the investors' objectives coincides with the growth optimality of their strategies. The equilibrium solutions are closely connected to the asymptotic behavior of the evolutionary model.

Evolutionary models of portfolio selection are related to the literature on growth optimal portfolios, see e.g. Hakansson (1970), Thorp (1971), Algoet and Cover (1988), Cover (1991), Hakansson and Ziemba (1995), Browne and Whitt (1996), Karatzas and Shreve (1998), and Aurell et al. (2000). As common in mathematical finance and in contrast to the evolutionary approach, these models usually assume an exogenous price process. Equilibrium consequences are neglected in these models. The current model makes a connection between an evolutionary approach and continuous time processes which are commonly used in mathematical finance. This has two implication. Techniques from stochastic analysis can be used for the investigation of the proposed model. At the same time, equilibrium effects are treated endogenously.

The balance of this chapter is organized as follows. In Section 4.2 we present the discrete time model of dynamic asset allocation of Blume and Easley (1992). In Section 4.3 we provide a continuous time approximation of the wealth process and suggest an economically meaningful model for the dividend processes. In Section 4.4 we study the long-run asymptotic behavior of the continuous time approximation of the wealth process in the deterministic case and examine a rational benchmark. Section 4.5 concludes. Auxiliary results are presented in an appendix.

## 4.2 Modeling Dynamic Asset Allocation

### 4.2.1 The Economy

In this section we provide a model of dynamic portfolio allocation and the evolution of wealth of investors in a financial market. By $i \in I = \{1, 2, \ldots, I\}$ we denote a finite set of investors who can invest into assets $k \in K = \{1, 2, \ldots, K\}$ at discrete points in time $t \in \mathbb{N}$.

At time $t$, investor $i \in I$ is endowed with wealth $w_i^t \in \mathbb{R}_+$. For the vector of agents' wealth we will write $w^t = (w_i^t)_{i \in I}$. At each point in time $t$ each investor $i$ acquires a portfolio $a_i^t = (a_{i,1}^t, a_{i,2}^t, \ldots, a_{i,K}^t)$; here $a_{i,k}^t$ denotes the number of shares of asset $k$ in the portfolio. For simplicity, we assume that assets live only for one period and are re-born at every period. Denoting the price of one share of asset $k$ by $\rho_k^t$, the $I$ budget constraints of the investors $i \in I$ can be written in the following form:

$$w_i^t = \sum_{k=1}^{K} \rho_k^t \cdot a_{i,k}^t \tag{4.1}$$

The prices are determined in a Walrasian market by the $K$ equilibrium equations

$$\bar{a}_k^t = \sum_{i=1}^{I} a_{i,k}^t \tag{4.2}$$

where $\bar{a}_k^t > 0$ is the total supply of asset $k$ in period $t$. For simplicity, we suppose that the supply of each asset does not depend on time and is non-random. By an appropriate renormalization of the payoffs of the assets we may and will assume that $\bar{a}_k^t \equiv 1$ for all $k \in K$. Economically, this hypothesis could be expressed in terms of a stock split. The *budget shares* of the assets in the portfolio of the investors are given by

$$\lambda_{i,k}^t = \frac{\rho_k^t \cdot a_{i,k}^t}{w_i^t} \tag{4.3}$$

The sequence of budget shares $\lambda_i = (\lambda_i^t)_{t\in\mathbb{N}} = (\lambda_{i,1}^t, \lambda_{i,2}^t, \ldots, \lambda_{i,K}^t)_{t\in\mathbb{N}}$ will be called the *trading strategy* of investor $i$.

Rewriting (4.2), we obtain the following equation for the market-clearing price:

$$\rho_k^t = \sum_{i=1}^{I} \lambda_{i,k}^t \cdot w_i^t \tag{4.4}$$

The shares bought at time $t$ pay a dividend at time $t+1$ which we will assume to be random. We let $(\Omega, \mathcal{F}, P)$ be a probability space. By $A_k^{t+1} : \Omega \to \mathbb{R}_+$ we denote the dividend payment of asset $k$ at time $t+1$. We will assume that all random quantities under consideration, i.e. $A_k^t$, $a_{i,k}^t$, and $w_i^t$ ($i \in I$, $k \in K$, $t \in \mathbb{N}$), are measurable.

Total dividend payments received by agents $i$ at time $t+1$ can be calculated as

$$D_i^{t+1} = \sum_k a_{i,k}^t \cdot A_k^{t+1} \tag{4.5}$$

The quantities we considered so far were given by their nominal value. The real wealth of any investor must be described as a fraction of total wealth times the real value of the economy. To keep the analysis simple, we will abstract from growth and assume that the real value of the economy is constant over time and equal to 1. Hence, in real terms economic quantities are given by choosing total market wealth as numeraire. Real wealth of investor $i$ at time $t$ is thus given by relative wealth

$$r_i^t = \frac{w_i^t}{\sum_{j=1}^I w_j^t} \tag{4.6}$$

Normalizing the prices of the assets by the market wealth we obtain the real prices of asset $k$ at date $t$:

$$q_k^t = \frac{\rho_k^t}{\sum_{i=1}^I w_i^t} = \sum_{i=1}^{I} \lambda_{i,k}^t r_i^t \tag{4.7}$$

The real payoff of asset $k$ at time $t+1$ can be calculated as

$$R_k^{t+1} = \frac{A_k^{t+1}}{\sum_{l=1}^K A_l^{t+1}} \tag{4.8}$$

### 4.2.2 The Wealth Dynamics in Discrete Time

Apart from the choice of the investments and the market structure, we have to describe how the wealth of the investors is determined in period $t+1$. We investigate the case of investors who never consume, but reinvest their investment earnings completely. For simplicity, we assume that investors do not receive income from labor. Hence, we suppose that $w_i^{t+1} = D_i^{t+1}$ for all times $t$ and agents $i$. We may rewrite the evolution of relative wealth as

$$
\begin{aligned}
r_i^{t+1} &= r_i^t \sum_k \frac{\lambda_{i,k}^t}{\sum_j \lambda_{j,k}^t r_j^t} R_k^{t+1} \\
&= r_i^t + r_i^t \left( \sum_{k=1}^{K} R_k^{t+1} \frac{\lambda_{i,k}^t}{\sum_{j=1}^{I} \lambda_{j,k} r_j^t} - 1 \right).
\end{aligned}
\tag{4.9}
$$

We will study the case in which the trading strategies $\lambda_{i,k}^t = \lambda_{i,k}$ do not depend on time. Hence, we will drop the index $t$. In this case, the wealth dynamic is only triggered by the random payments. We will always stick to the following assumption.

***Assumption 4.2.1.*** *All agents invest a strictly positive amount into any asset, i.e. the values $\lambda_{i,k}$ are strictly positive. In economic terms, all agents are completely diversified.*

## 4.3 The Wealth Dynamics in Continuous Time

### 4.3.1 A Continuous Time Approximation

We will now provide a continuous time approximation of the model assuming that dividends are paid at a higher frequency. It is shown that the discrete time model converges to an integral equation in a random environment.

Given $n \in \mathbb{N}$, we let a new time grid be given by the time points $\{l \cdot n^{-1} : \ l \in \mathbb{N}_0\}$. Dividends are paid at these dates, and the corresponding dividend process is a discrete time stochastic process denoted by $(A^{(n),s/n})_{s\in\mathbb{N}_0}$. By convention, we fix $A^{(n),0} = a_0 \in \mathbb{R}_+^K$.

***Assumption 4.3.1.*** *For all $n \in \mathbb{N}$ and $s \in \mathbb{N}_0$, we suppose that with probability one*

$$\sum_{k=1}^{K} A_k^{(n),s/n} > 0.$$

Thus, the real returns of the assets are given by the expressions

$$R_k^{(n),s/n} = \frac{A_k^{(n),s/n}}{\sum_{l=1}^{K} A_l^{(n),s/n}}. \tag{4.10}$$

As before we suppose that trading takes place immediately after dividends have been received, but we will no longer assume that total wealth is invested. At times $0, \frac{1}{n}, \frac{2}{n}, \ldots$ agents invest only a fraction $\alpha^n \in (0, 1]$ of their wealth in the market. This assumption modifies the dynamics described by equation (4.9). For fixed $n$, the wealth dynamics is now given by the following recursive scheme

$$r_i^{(n)}(t_{n,l+1}) = (1-\alpha^n)\cdot r_i^{(n)}(t_{n,l}) + \alpha^n \cdot r_i^{(n)}(t_{n,l}) \sum_{k=1}^{K} \frac{\lambda_{i,k} R_k^{(n),t_{n,l+1}}}{\sum_{j=1}^{I} r_j^{(n)}(t_{n,l})\lambda_{j,k}}\,, \tag{4.11}$$

where $t_{n,l} = \frac{l}{n}$ and $r_0^{(n)} = r_0 \in \Delta_I$. Here, $\Delta_I$ denotes the simplex in $\mathbb{R}^I$.

We are interested in a continuous time approximation for $n \to \infty$ where we choose $\alpha^n = \frac{1}{n}$. For this purpose, it is convenient to extend all discrete time processes to continuous time. The continuous time extension of real returns $R^{(n)}$ is defined by the piecewise constant process

$$R^{(n)} := R^{(n),0}\cdot \mathbf{1}_{\{0\}} \;+\; \sum_{s=0}^{\infty} R^{(n),(s+1)/n}\cdot \mathbf{1}_{\left(\frac{s}{n},\frac{s+1}{n}\right]}. \tag{4.12}$$

The wealth process $r^{(n)}$ is extended to continuous time by linear interpolation. For $t_{n,l} \le s \le t_{n,l+1}$ and $i = 1, 2, \ldots, I$, we let

$$\begin{aligned} r_i^{(n)}(s) \;&:=\; r_i^{(n)}(t_{n,l}) + n(s-t_{n,l})\left(r_i^{(n)}(t_{n,l+1}) - r_i^{(n)}(t_{n,l})\right) \\ &=\; r_i^{(n)}(t_{n,l}) \\ &\quad +\; \int_{t_{n,l}}^{s} r_i^{(n)}(t_{n,l})\left(\sum_{k=1}^{K} \frac{\lambda_{i,k} R_k^{(n),t_{n,l+1}}}{\sum_{j=1}^{I} r_j^{(n)}(t_{n,l})\lambda_{j,k}} \;-\; 1\right) du \end{aligned} \tag{4.13}$$

As preparation for the approximation theorem, we need the following proposition. Again, $\Delta_I$ and $\Delta_K$ denote the simplices in $\mathbb{R}^I$ and $\mathbb{R}^K$, respectively.

***Proposition 4.3.2.*** *Let $T : \mathbb{R}_+ \to \Delta_K$ be measurable. Assume that $r_0 \in \Delta_I$. Then the coupled integral equations*

$$r_i(s) \;=\; r_{i,0} + \int_0^s r_i(s')\left(\sum_{k=1}^{K} \frac{\lambda_{i,k} T_k^{s'}}{\sum_{j=1}^{I} r_j(s')\lambda_{j,k}} - 1\right) ds', \tag{4.14}$$

*with $i = 1, 2, \ldots, I$, possess a unique continuous solution $r : \mathbb{R}_+ \to \Delta_I$.*

*Proof.* Since all norms on finite dimensional vector spaces are equivalent, we do not have to specify a particular norm on $\mathbb{R}^I$ and $\mathbb{R}^K$, respectively. Of course, bounds and Lipschitz constants depend on the choice of the norms. For simplicity, we will denote the norms by $\|\cdot\|$.

The right hand side of the integral equation (4.14) depends on a function $\psi$ with domain $\Delta_I \times \Delta_K$ defined by

$$\psi_i(r,T) = r_i\left(\sum_{k=1}^{K} \frac{\lambda_{i,k}T_k}{\sum_{j=1}^{I} r_j\lambda_{j,k}} - 1\right). \tag{4.15}$$

$\psi$ is both bounded by some constant $B$ and Lipschitz continuous with constant $L$ as can be seen by the following arguments. First, $\psi$ is affine in $T$. Second, observe that by assumption the trading strategies $\lambda_{j,k}$ are strictly positive. Hence, the zeros of the linear mapping

$$r \mapsto \sum_{j=1}^{I} r_j\lambda_{j,k}$$

are not contained in $\Delta_I$. It follows that $\psi$ is continuously differentiable on its compact domain, hence both bounded and Lipschitz continuous.

We first verify uniqueness. If $r_1$ and $r_2$ are two solutions, then by Lipschitz continuity of $\psi$ we obtain

$$\sup_{0\le s\le t} \|r_1(s) - r_2(s)\| \le L\, t \sup_{0\le s\le t} \|r_1(s) - r_2(s)\|.$$

This implies uniqueness for $t < 1/L$. A concatenation argument implies the identity $r_1(s) = r_2(s)$ for any $s \in \mathbb{R}_+$.

Next we prove existence. Define a sequence of functions $\rho^{(n)} : \mathbb{R}_+ \to \mathbb{R}^I$ by the following recursive scheme

$$\rho^{(n)}(t) \quad = \quad r_0 + \int_0^t \sum_{l=0}^{\infty} 1_{[\tau_{n,l},\tau_{n,l+1})}(u)\, \psi\left(\rho^{(n)}(\tau_{n,l}), \mathcal{T}_n^u\right) du\,,$$

where $\tau_{n,l} = l\,2^{-n}$, $n \in \mathbb{N}$, $l \in \mathbb{N}_0$. Here, the second argument of $\psi$ equals the average

$$\mathcal{T}_n^u := 2^n \cdot \int_{\tau_{n,l}}^{\tau_{n,l+1}} T^u du \qquad (u \in [\tau_{n,l}, \tau_{n,l+1})).$$

$\rho^{(n)}$ is continuous with values in $\Delta_I$. As $\psi$ is uniformly bounded, we obtain $\|\rho^{(n)}(t) - \rho^{(n)}(s)\| \le B|t-s|$. Thus $K = \{\rho^{(n)} :\ n \in \mathbb{N}\}$ is relatively compact by the Theorem of Arćela-Ascoli. Hence, there exists a continuous function $r : [0,\infty) \to \Delta_I$ and a subsequence $\rho^{(n')}$ converging to $r$ uniformly on compacts.

We show that $r$ is a solution of the integral equation. We need to verify that for all $t \ge 0$

$$\lim_{n'\to\infty} \int_0^t \sum_{l=0}^{\infty} 1_{[\tau_{n',l},\tau_{n',l+1})}(u)\psi\left(\rho^{(n')}(\tau_{n',l}), \mathcal{T}_{n'}^u\right) du \quad = \quad \int_0^t \psi\left(r(u), T^u\right) du\,.$$

We obtain by the triangle inequality and Lipschitz continuity,

$$
\begin{aligned}
&\left\| \int_0^t \sum_{l=0}^{\infty} 1_{[\tau_{n',l},\tau_{n',l+1})}(u)\, \psi\left(\rho^{(n')}(\tau_{n',l}), \mathcal{T}^u_{n'}\right) du \;-\; \int_0^t \psi\left(r(u), T^u\right) du \right\| \\
\leq\; & \int_0^t \left\| \sum_{l=0}^{\infty} 1_{[\tau_{n',l},\tau_{n',l+1})}(u)\, \psi\left(\rho^{(n')}(\tau_{n',l}), T^u\right) - \psi\left(r(u), T^u\right) \right\| du \\
& + \int_0^t \sum_{l=0}^{\infty} 1_{[\tau_{n',l},\tau_{n',l+1})}(u) \left\| \psi\left(\rho^{(n')}(\tau_{n',l}), \mathcal{T}^u_{n'}\right) - \psi\left(\rho^{(n')}(\tau_{n',l}), T^u\right) \right\| du \\
\leq\; & Lt \cdot \max_{0 \leq \tau_{n',l} \leq t} \; \sup_{\tau_{n',l} \leq u \leq \tau_{n',l+1}} \left\| \rho^{n'}(\tau_{n',l}) - r(u) \right\| + L \int_0^t \left\| \mathcal{T}^u_{n'} - T^u \right\| du.
\end{aligned}
$$

$r$ is uniformly continuous on compact sets. Thus, the first term converges to 0 by choice of $\rho^{(n')}$. The second term converges to 0, since the averages $\mathcal{T}_{n'}$ converge to $T$ in $L^1([0,t])$ for any $t > 0$.[1] □

The continuous time approximation of the evolutionary stock market model of Blume and Easley (1992) is a consequence of the following theorem.

***Theorem 4.3.3.*** *Let $(\Omega, \mathcal{F}, P)$ be a probability space.*

*For each $n \in \mathbb{N}$, we let $(R^{(n),(s-1)/n})_{s\in\mathbb{N}}$ be a sequence of random variables on $\Omega$ with values in $\Delta_K$. $R^{(n)}$ is extended to a continuous time process by (4.12). Assume that $r^{(n)}$ is defined according to (4.11) and (4.13) with $r^{(n)}(0) = r_0 \in \Delta_I$.*

*Let $(T^s)_{s\in\mathbb{R}_+}$ be a stochastic process on $\Omega$ with values in $\Delta_K$ that is jointly measurable in $\omega \in \Omega$ and $s \in \mathbb{R}_+$. Suppose that $r$ is the pathwise unique continuous solution of (4.14).*

*Then there exists for every $t \geq 0$ a non-random constant $D$ such that for all $n \in \mathbb{N}$ the following inequality holds:*

$$
\sup_{0 \leq s \leq t} \|r(s) - r^{(n)}(s)\|_{\mathbb{R}^I} \leq D \cdot \left( \frac{1}{n} + \int_0^{t+\frac{1}{n}} \|T^u - R^{(n),u}\|_{\mathbb{R}^K} \, du \right), \tag{4.16}
$$

*where $\|\cdot\|_{\mathbb{R}^I}$ and $\|\cdot\|_{\mathbb{R}^K}$ are given norms on $\mathbb{R}^I$ and $\mathbb{R}^K$, respectively.*

*Proof.* The proof of the consistency of the Euler scheme can be divided into two steps. First, control the approximation error locally, and then find bounds on the global approximation error.

[1] The $L^1$-convergence of the averages can be verified by Doob's martingale convergence theorem. See e.g. the proof of Proposition 4.3.7.

We choose the Lipschitz constant $L$ and the bound $B$ as in the proof of Proposition 4.3.2. For simplicity, we omit the index $n$ from $t_{n,k}$. We define and bound a local approximation error $l_{n,k}$ as follows:

$$\begin{aligned} l_{n,k} &:= \sup_{t_k \le s \le t_{k+1}} \left\| \int_{t_k}^{s} \psi(r(u), T^u)\, du - \int_{t_k}^{s} \psi(r(t_k), R^{(n),t_{k+1}})\, du \right\| \\ &\le L \cdot \left\{ \int_{t_k}^{t_{k+1}} \|r(u) - r(t_k)\| du + \int_{t_k}^{t_{k+1}} \|T^u - R^{(n),t_{k+1}}\| du \right\} \qquad (4.17) \end{aligned}$$

Here, we used the Lipschitz continuity of $\psi$. Now observe that (4.14) implies

$$\|r(u) - r(t_k)\| \le \int_{t_k}^{u} \|\psi(r(v), T^v)\|\, dv \le B(u - t_k).$$

We get therefore for the right hand side of (4.17) an upper bound

$$\begin{aligned} & L \cdot \left\{ \int_{t_k}^{t_{k+1}} B(u - t_k)\, du + \int_{t_k}^{t_{k+1}} \|T^u - R^{(n),t_{k+1}}\| du \right\} \\ \le\ & L \cdot \left\{ \frac{B}{2n^2} + \int_{t_k}^{t_{k+1}} \|T^u - R^{(n),t_{k+1}}\| du \right\} \qquad (4.18) \end{aligned}$$

Observe that for $t_k \le s \le t_{k+1}$ we can rewrite (4.13)

$$r^{(n)}(s) = r^{(n)}(t_k) + \int_{t_k}^{s} \psi(r^{(n)}(t_k), R^{(n),t_{k+1}})\, du \qquad (4.19)$$

Next, we define the error

$$\delta_k = \delta_k^{(n)} = \|r(t_k) - r^{(n)}(t_k)\| \qquad (4.20)$$

and observe that by (4.14) and (4.19) for every $t_k \le s \le t_{k+1}$

$$\begin{aligned} \|r(s) - r^{(n)}(s)\| &\le \|r(t_k) - r^{(n)}(t_k)\| \\ &\quad + \left\| \int_{t_k}^{s} \psi(r(u), T^u) - \psi(r^{(n)}(t_k), R^{(n),t_{k+1}})\, du \right\| \\ &\le \delta_k + \left\| \int_{t_k}^{s} \psi(r(t_k), R^{(n),t_{k+1}}) - \psi(r^{(n)}(t_k), R^{(n),t_{k+1}})\, du \right\| \\ &\quad + \left\| \int_{t_k}^{s} \psi(r(u), T^u) - \psi(r(t_k), R^{(n),t_{k+1}})\, du \right\| \\ &\le \delta_k \left( 1 + L\frac{1}{n} \right) + l_{n,k}. \qquad (4.21) \end{aligned}$$

In particular, taking $s = t_{k+1}$ we get

$$\delta_{k+1} \leq \delta_k \left(1 + L\frac{1}{n}\right) + l_{n,k} \tag{4.22}$$

Observing $\delta_0 = 0$, we derive for $0 \leq k \leq \lfloor tn \rfloor + 1$ by induction

$$\delta_k \ \leq \ \left(1 + \frac{L}{n}\right)^k \sum_{m=0}^{k-1} l_{n,m} \tag{4.23}$$

Hence, we can estimate for $0 \leq k \leq \lfloor tn \rfloor + 1$

$$\begin{aligned}
\delta_k \ &\leq \ \left(1 + \frac{L}{n}\right)^{\lfloor tn \rfloor + 1} \sum_{k=0}^{\lfloor tn \rfloor} l_{n,k} \\
&\leq \ \left(1 + \frac{L}{n}\right)^{\lfloor tn \rfloor + 1} \cdot L \cdot \sum_{k=0}^{\lfloor tn \rfloor} \left( \frac{B}{2n^2} + \int_{t_k}^{t_{k+1}} \|T^u - R^{(n),t_{k+1}}\| du \right) \\
&\leq \ D' \cdot \left( \frac{1}{n} + \int_0^{t+\frac{1}{n}} \|T^u - R^{(n),u}\| du \right)
\end{aligned} \tag{4.24}$$

where $D'$ is some constant depending only on $t$. This together with (4.21) implies that

$$\sup_{0 \leq s \leq t} \|r(s) - r^{(n)}(s)\| \leq D \cdot \left( \frac{1}{n} + \int_0^{t+\frac{1}{n}} \|T^u - R^{(n),u}\| du \right), \tag{4.25}$$

where $D$ is some constant depending only on $t$. □

The following corollary summarizes the final result of the current section, i.e. the continuous time approximation of the evolutionary model.

***Corollary 4.3.4.*** *Consider the same setting as in Theorem 4.3.3. Let*

$$Y_t^n := \int_0^t \|T^u - R^{(n),u}\|_{\mathbb{R}^K} \, du \ . \tag{4.26}$$

*Then the following implications hold:*

*(1) If $Y_t^n$ converges for all $t \in \mathbb{R}_+$ to $0$ almost surely, then $r^{(n)}$ converges to $r$ uniformly on compacts with probability $1$.*

*(2) If $Y_t^n$ converges for all $t \in \mathbb{R}_+$ to $0$ in probability, then $r^{(n)}$ converges to $r$ uniformly on compacts in probability and in $L^p$ for $p \in [1, \infty)$.*

*(3) If $Y_t^n$ converges for all $t \in \mathbb{R}_+$ to $0$ in $L^\infty$, then $r^{(n)}$ converges to $r$ uniformly on compacts in probability and in $L^p$ for $p \in [1, \infty]$.*

*Proof.* Inequality (4.16) implies clearly the convergence of $r^{(n)}$ to $r$ given appropriate conditions on the convergence of $Y_t^n$ to 0 for any $t \in \mathbb{R}_+$. Since all quantities we are dealing with are uniformly bounded, convergence in any $L^p$-norm ($p \in [1, \infty)$) and convergence in probability are equivalent. □

### 4.3.2 Dividend Processes in Continuous Time

The limiting process $T$ in Theorem 4.3.3 and Corollary 4.3.4 can be interpreted as the real return process in continuous time. As a limiting process, $T$ is specified by the real returns defined in (4.10) and thus by the family of discrete time dividend processes $(A^{(n),s/n})_{s \in \mathbb{N}_0}$ ($n \in \mathbb{N}$). In this section we provide an economically meaningful model for these dividend processes and investigate the convergence of the real returns to a continuous time process $T$.

We fix a probability space $(\Omega, \mathcal{F}, P)$, and assume that all random variables and processes are defined on $\Omega$. We suppose now that the dividend streams are driven by value processes earned by firms. More specifically, let $S^t \in \mathbb{R}_+^K$ be the stochastic process of the excess value generated by $K$ firms corresponding to the assets $k = 1, 2, \dots, K$, i.e. the process of cumulated dividends. We make the following assumption:

***Assumption 4.3.5.*** *The value process $S^t$ is cadlag and strictly increasing in the following sense: for given $t \in \mathbb{R}_+$, $\omega \in \Omega$ and $\epsilon > 0$ it holds that*

- $\forall k$: $S_k^{t+\epsilon}(\omega) - S_k^t(\omega) \geq 0$
- $\exists k$: $S_k^{t+\epsilon}(\omega) - S_k^t(\omega) > 0$

We will assume that the value process $S^t$ is related to the dividend payments in the following way:

(1) $A^0 = S^0$,

(2) $A^t = S^t - S^{t-1}$ for $t \in \mathbb{N}$.

In other words, at time $t$ the firm pays the complete incremental value generated between times $t-1$ and $t$ as dividends to the investors. The real payoff of asset $k$ at time $t+1$ can therefore be calculated as

$$R_k^{t+1} = \frac{A_k^{t+1}}{\sum_{l=1}^K A_l^{t+1}} = \frac{S_k^{t+1} - S_k^t}{\sum_{l=1}^K (S_l^{t+1} - S_l^t)}. \tag{4.27}$$

In the continuous time approximation, dividends are paid at a higher frequency. Given $n \in \mathbb{N}$, we define on the new time grid $\frac{1}{n}\mathbb{N}$ for $s \in \mathbb{N}_0$

(1) $A^{(n),0} = S^0$,

(2) $A^{(n),(s+1)/n} = S^{(s+1)/n} - S^{s/n}$.

In terms of the value process, real returns are thus given by

$$R_k^{(n),(s+1)/n} = \frac{A_k^{(n),(s+1)/n}}{\sum_{l=1}^K A_l^{(n),(s+1)/n}} = \frac{S_k^{(s+1)/n} - S_k^{s/n}}{\sum_{l=1}^K (S_l^{(s+1)/n} - S_l^{s/n})}. \tag{4.28}$$

At time 0, we obtain returns not depending on $n$:

$$R_k^{(n),0} = \frac{A_k^{(n),0}}{\sum_{l=1}^K A_l^{(n),0}} = \frac{S_k^0}{\sum_{l=1}^K S_l^0}. \tag{4.29}$$

We extend again $R^{(n)}$ to a continuous time process $(R^{(n)}(\omega, u))_{u\geq 0}$ by formula (4.12). If the real dividends $R^{(n)}$ converge in an appropriate sense to a limiting process $T$, we can apply Theorem 4.3.3 and Corollary 4.3.4 to obtain a continuous time approximation of the wealth process. We are thus interested in the question when the stochastic processes $R^{(n)}$ converge to a limiting process $T$ and how this process is related to the firms' value process $S$. For this purpose, it is very helpful to establish a representation of $S$ is terms of locally finite kernels.

**A representation of the firm value process.** By Assumption 4.3.5, for $\omega \in \Omega$ the components $S_k(\omega)$ $(k = 1, \dots, K)$ of the firm value process are cumulative distribution functions of a positive locally finite Borel measure $\mu_k(\omega)$ on $\mathbb{R}_+$. More precisely, $\mu_k$ $(k = 1, 2, \dots, K)$ is a locally finite kernel from $\Omega$ to $\mathbb{R}_+$. Here, a mapping $\mu : \Omega \times \mathcal{B}(\mathbb{R}_+) \to \bar{\mathbb{R}}_+$ is called a locally finite kernel, if

(1) $\mu(\cdot, B) : \Omega \mapsto \bar{\mathbb{R}}_+$ is measurable;

(2) $\mu(\omega, \cdot)$ is a locally finite measure on $\mathbb{R}_+$ for all $\omega \in \Omega$.

Given a probability measure $P$ on $(\Omega, \mathcal{F})$, every locally finite kernel $\mu$ from $\Omega$ to $\mathbb{R}_+$ induces a unique $\sigma$-finite measure $P\mu$ on $(\Omega \times \mathbb{R}_+, \mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+))$, cf. Proposition 4.6.2. $P\mu$ is uniquely defined by setting

$$P\mu(A \times B) := \int_A \mu(\omega, B)\, dP \qquad (A \in \mathcal{F},\ B \in \mathcal{B}(\mathbb{R}_+)). \tag{4.30}$$

The following theorem provides a canonical representation of the firm value process $S$ which is useful when investigating the convergence to a continuous time dividend process. The notion of *exhausting sequence* is given in Definition 4.6.1 in the appendix.

***Theorem 4.3.6.*** *Suppose that Assumption 4.3.5 holds. Then we find a canonical representation of $S^t$ in terms of a locally finite kernel $\mu$ from $\Omega$ to $\mathbb{R}^+$ and measurable functions $f_k : \Omega \times \mathbb{R}_+ \to \mathbb{R}_+$. Namely, for every $1 \leq k \leq K$ and for every $t \geq 0$ the firm value process $S : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}^K$ satisfies for $P$-almost all $\omega \in \Omega$,*

$$S_k^t(\omega) = \int \mathbf{1}_{[0,t]}(u)\ f_k(\omega, u)\ \mu(\omega, du). \tag{4.31}$$

*Furthermore, the sum of the functions $(f_k(\omega, u))_{k=1,\ldots,K}$ is $P\mu$-almost everywhere positive. For every exhausting sequence $(C_N)$ for $P$ and $\mu$, the functions $f 1_{C_N}$ are $P\mu$-integrable .*

*Proof.* Let $\mu_k$ be the measure associated with the cumulative distribution function $S_k$. Define $\mu := \sum_{k=1}^K \mu_k$. By Assumption 4.3.5 $\mu_k(\omega, \cdot)$ is a locally finite measure on $\mathbb{R}_+$. The mapping $\mu_k(\cdot, B) : \Omega \mapsto \bar{\mathbb{R}}_+$ is measurable. The same is true for $\mu$.

By Lemma 4.6.2 both $P\mu$ and $P\mu_k$ are $\sigma$-finite measures. Moreover, $P\mu$ dominates $P\mu_k$. Thus, by the theorem of Radon-Nikodym, there exist densities $f_k : \Omega \times \mathbb{R}_+ \to \mathbb{R}_+$ such that $d(P\mu_k) = f_k d(P\mu)$ $(k = 1. \ldots, K)$.

For any $F \in \mathcal{F}$ we obtain

$$\int_F S_k^t(\omega)\, P(d\omega) = P\mu_k(F \times [0, t]) = \int_F \int 1_{[0,t]}(u) f(\omega, u)\, d\mu(\omega, u)\, P(d\omega)\,.$$

Since the equality holds for all $F \in \mathcal{F}$, we obtain (4.31).

Observe that $dP\mu = \sum_{k=1}^K dP\mu_k = \left(\sum_{k=1}^K f_k\right) dP\mu$. Hence, we can conclude that $\sum_{k=1}^K f_k = 1$ $P\mu$-almost everywhere.

Finally, let $(C_N)$ be an exhausting sequence for $P$ and $\mu$. This implies that

$$\int 1_{C_N} f_k\, dP\mu = P\mu_k(C_N) \leq P\mu(C_N) < \infty.$$

Thus, the functions $1_{C_N} f_k$ are $P\mu$-integrable. $\square$

**Convergence to a continuous time dividend process.** We now return to the question when $R^{(n)}$ converges to a process $T$. The proof of the following proposition is based on a martingale argument.

***Proposition 4.3.7.*** *Suppose that Assumption 4.3.5 holds. We suppose that $S^t$ is represented according to (4.31). Then, for $P\mu$-almost all $(\omega, u)$, the limit of $R^{(n)}(\omega, u)$ exists for $n \to \infty$ and equals*

$$\lim_{n\to\infty} R_k^{(n)}(\omega, u) = \frac{f_k(\omega, u)}{\sum_{l=1}^K f_l(\omega, u)}. \tag{4.32}$$

*Proof.* By Lemma 4.6.2 we can find an exhausting sequence $(C_N)_N$ for $P$ and $\mu$. It clearly suffices to verify the claim for $P\mu$-almost every $(\omega, s) \in C_N$ and any $N \in \mathbb{N}$. By definition, $C_N = F_N \times [0, \alpha_N)$ for some $F_N \in \mathcal{F}$ and $\alpha_N > 0$.

W.l.o.g. suppose that $P\mu(C_N) > 0$. Since $P\mu(C_N)$ is finite, we may normalize $P\mu$. Thus, we assume w.l.o.g. that $P\mu \in \mathcal{M}_1(C_N)$.

By $\mathcal{F}_n$ we denote the $\sigma$-algebra on $\mathbb{R}_+$ generated by the partition

$$\{\{0\}\} \cup \left\{ \left( \frac{l-1}{n}, \frac{l}{n} \right], \; l \in \mathbb{N} \right\}.$$

$\mathcal{F}_n$ induces a $\sigma$-algebra $\mathcal{G}_n^N$ on $C_N$, namely

$$\mathcal{G}_n^N = \Big( F_N \cap \mathcal{F} \Big) \otimes \Big( [0, \alpha_N) \cap \mathcal{F}_n \Big),$$

where $F_N \cap \mathcal{F} = \{F_N \cap F : \; F \in \mathcal{F}\}$, $[0, \alpha_N) \cap \mathcal{F}_n = \{[0, \alpha_N) \cap E : \; E \in \mathcal{F}_n\}$, respectively.

Let $g : C_N \to \mathbb{R}$ be measurable and integrable with respect to $P\mu$. Doob's martingale convergence theorem for directed index sets implies that $E_{P\mu}(g|\mathcal{G}_n^N)$ converges $P\mu$-almost surely to $g$ as $n \to \infty$. This result can be applied to $1_{C_N} f_k$, since $E_{P\mu}(1_{C_N} f_k) < \infty$ by Theorem 4.3.6.

For $(\omega, s) \in C_N$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ the number $(\lfloor sn \rfloor + 1)/n$ is strictly smaller than $\alpha_N$. We obtain therefore for $n \geq n_0$

$$\begin{aligned}
R_k^{(n)}(\omega, s) &= \frac{S_k^{(\lfloor sn \rfloor)+1/n} - S_k^{\lfloor sn \rfloor/n}}{\sum_{l=1}^K \left( S_l^{(\lfloor sn \rfloor+1)/n} - S_l^{\lfloor sn \rfloor/n} \right)} \\
&= \left( \int \mathbf{1}_{(\lfloor sn \rfloor/n, (\lfloor sn \rfloor+1)/n]} f_k d\mu \right) \cdot \left( \sum_{l=1}^K \int \mathbf{1}_{(\lfloor sn \rfloor/n, (\lfloor sn \rfloor+1)/n]} f_l d\mu \right)^{-1} \\
&= \big( E_{P\mu}(f_k|\mathcal{G}_n^N)(\omega, s) \big) \cdot \left( \sum_{l=1}^K E_{P\mu}(f_l|\mathcal{G}_n^N)(\omega, s) \right)^{-1}
\end{aligned}$$

The last term converges $P\mu$-almost everywhere to $f_k(\omega, s) \cdot \left( \sum_{l=1}^K f_l(\omega, s) \right)^{-1}$. $\square$

**Dividend convergence and Euler approximation.** In this paragraph we provide sufficient conditions on the firms' value process $S^t$ which ensure the convergence of the discrete time wealth processes $r^{(n)}$ to a continuous time process $r$. In terms of the family of random variables $Y_t^n$ $(t \in \mathbb{R}_+, \; n \in \mathbb{N})$ conditions have been derived in Section 4.3.1, see in particular Corollary 4.3.4. We will now combine these results with representation (4.31) of Theorem 4.3.6.

If the measure $\mu(\omega,\cdot)$ dominates the Lebesgue measure for $P$-almost all $\omega \in \Omega$, strong implications can be derived. In this case, with probability 1 the limiting statement (4.32) holds both $\mu$- and Lebesgue-almost everywhere, and we obtain the following Euler approximation.

***Corollary 4.3.8.*** *Suppose that the Assumption 4.3.5 holds, and let a representation of the value process $S$ be given according to Theorem 4.3.6. Suppose that $P$-almost surely $\mu$ dominates the Lebesgue measure. For $k = 1, \dots, K$, we set $g_k = f_k$ if $\sum_{l=1}^K f_l > 0$, and $g_k = 1$ else. Define the process $T = g_k \cdot \left(\sum_{l=1}^K g_l\right)^{-1}$. Then $r^{(n)}$ converges to $r$ defined in (4.14) uniformly on compacts with probability 1.*

*Proof.* By Proposition 4.3.7 we obtain that $\lim_{n\to\infty} R^{(n)}(\omega, s) = T(\omega, s)$ $P\mu$-almost everywhere. Set $L \in \mathcal{F}\otimes\mathcal{B}(\mathbb{R}_+)$ be the set of all $(\omega, s)$ such that $\lim_{n\to\infty} R^{(n)}(\omega, s)$ exists and equals $T(\omega, s)$. Denote by $L^c$ the complement of $L$, and let $B_\omega = \{s : (\omega, s) \notin L\}$. Then,

$$0 = \int L^c \, dP\mu = \int \mu(\omega, B_\omega)\, P(dw).$$

Hence, $\mu(\omega, B_\omega) = 0$ for $P$-almost all $\omega \in \Omega$. Since $\mu(\omega,\cdot)$ dominates the Lebesgue measure for $P$-almost all $\omega \in \Omega$, we obtain that $\lambda(B_\omega) = 0$ for $P$-almost all $\omega \in \Omega$ where $\lambda$ is the Lebesgue measure. Thus, we conclude that $P$-almost surely for $t \in \mathbb{R}_+$,

$$\lim_{n\to\infty} Y_t^n = \lim_{n\to\infty} \int \|T^u - R^{(n),u}\| du = \int \lim_{n\to\infty} \|T^u - R^{(n),u}\| du = 0.$$

Interchanging limit and integral is justified by the dominated convergence theorem, since $P$-almost surely $T$ and $R^{(n)}$ ($n \in \mathbb{N}$) are bounded in $\Delta_K$. The result follows from Theorem 4.3.3. $\square$

The condition on Corollary 4.3.8 is not always satisfied. Given a value process $S$, we can in general not expect to find a representation (4.31) such that $\mu$ dominates the Lebesgue measure as the next example shows.

***Example 4.3.9.*** *For the construction of the counterexample we may w.l.o.g. focus on the deterministic case. Let $K = 1$, and define $\nu := \sum_{l\in\mathbb{N}} \frac{1}{2^l}\delta_{q_l}$, where $q : \mathbb{N} \to \mathbb{Q}$, $l \mapsto q_l$ is a bijection. Assume that $S : \mathbb{R}_+ \to \mathbb{R}_+$ is given by $S^t = \nu([0, t])$. It is not possible to find a measure $\mu$ dominating the Lebesgue measure and a density $f$ which is $\mu$-almost surely positive such that $S$ can be represented by $S^t = \int \mathbf{1}_{[0,t]} f d\mu$ .*

In terms of the representing kernel $\mu$ in (4.31), Corollary 4.3.8 provides a sufficient condition on the firms' value process $S^t$ which ensures the convergence of the discrete time wealth processes $r^{(n)}$ to a continuous time process $r$. The next proposition and

Corollary 4.3.11 give a condition in terms of the representing densities $f_k$ $(k = 1, \dots, K)$. If these functions are sufficiently regular, then the continuous time approximation of the wealth process is valid – irrespectively of the properties of the representing kernel $\mu$.

***Proposition 4.3.10.*** *Suppose that Assumption 4.3.5 holds. Assume that there exists a canonical representation according to Theorem 4.3.6 such that the mappings $f_k(\omega, \cdot) : \mathbb{R}_+ \mapsto \mathbb{R}_+$ are Lebesgue-almost everywhere continuous for $1 \leq k \leq K$ with probability one. Then for $P$-almost all $\omega \in \Omega$ the sum of the functions $(f_k(\omega))_{k=1,\dots,K}$ is Lebesgue-almost everywhere positive and the limit of $R^{(n)}(\omega)$ exists Lebesgue-almost everywhere and equals*

$$\lim_{n\to\infty} R_k^{(n)}(\omega) = \frac{f_k(\omega)}{\sum_{l=1}^K f_l(\omega)}. \tag{4.33}$$

*Proof.* Observe that

$$0 = P\mu\left(\sum_{k=1}^K f_k(\omega, u) = 0\right) = \int \mu\left(\omega, \left\{s : \sum_{k=1}^K f_k(\omega, s) = 0\right\}\right) P(d\omega).$$

Thus, the sum of the functions $(f_k(\omega))_{k=1,\dots,K}$ is $\mu(\omega, \cdot)$-almost surely positive for $P$-almost all $\omega \in \Omega$.

Assumption 4.3.5 implies that the complement of any $\mu(\omega, \cdot)$-nullset lies densely in $\mathbb{R}_+$. The regularity of $f$ implies then the positivity of the sum of the components Lebesgue-almost everywhere with probability one.

Let $B_n(s) := \left[\frac{\lfloor sn \rfloor - 1}{n}, \frac{\lfloor sn \rfloor + 1}{n}\right]$. Note that $\mu(t, t+\epsilon) > 0$ for $t \in \mathbb{R}_+$ and $\epsilon > 0$. Thus, by definition of $R^{(n)}$ it holds that

$$\frac{\inf_{u\in B_n(s)} f_k(u)}{\sum_{l=1}^K \sup_{u\in B_n(s)} f_l(u)} \leq R_k^{(n)}(s) \leq \frac{\sup_{u\in B_n(s)} f_k(u)}{\sum_{l=1}^K \inf_{u\in B_n(s)} f_l(u)}.$$

Since $f$ is Lebesgue-almost everywhere continuous, the claim follows. □

***Corollary 4.3.11.*** *Suppose that the assumptions of Proposition 4.3.10 are satisfied. For $k = 1, \dots, K$, we set $g_k = f_k$ if $\sum_{l=1}^K f_l > 0$, and $g_k = 1$ else. Define the process $T = g_k \cdot \left(\sum_{l=1}^K g_l\right)^{-1}$. Then $r^{(n)}$ converges to $r$ defined in (4.14) uniformly on compacts with probability* 1.

*Proof.* The proof is analogous to the last part of the proof of Corollary 4.3.8. □

## 4.4 Deterministic Dynamics

The continuous time wealth dynamics (4.14) is driven by the relative dividend process $T$. We are interested in the relative performance of the strategies which is characterized by the asymptotic behavior of the wealth process as $t \to \infty$. In this section we will focus on a special case – assuming that $T$ is deterministic and constant which corresponds to no dividend risk. While fundamentals are fixed, prices and wealth vary due to market interaction.

The wealth dynamics in the absence of fundamental risk is described by an autonomous differential equation. We will analyze its asymptotics employing the technique of Lyapunov functions. The analysis forms also the basis for the investigation of the more complex situation with a stochastic dividend process. This case will be investigated in Buchmann and Weber (2004b).

### 4.4.1 The Semiflow of the Wealth Dynamics

We suppose $T \equiv \pi$ for fixed $\pi \in \Delta_K$. For the whole section we make the following assumption.

***Assumption 4.4.1.*** *The real dividends $\pi$ are strictly positive, i.e. $\pi_k > 0$ for $1 \leq k \leq K$.*

Define a mapping $N : \Delta_I \to \mathbb{R}^I$ by

$$N_i(r) \quad = \quad \sum_{k=1}^{K} \frac{\pi_k \lambda_{i,k}}{\sum_{j=1}^{I} r_j \lambda_{j,k}} - 1. \tag{4.34}$$

Moreover, let the vector field $\psi : \Delta_I \to \mathbb{R}^I$ be given by

$$\psi_i(r) \quad = \quad r_i \cdot N_i(r). \tag{4.35}$$

Then the integral equation (4.14) reduces to an autonomous differential equation, namely

$$\dot{r}(t) \quad = \quad \psi(r), \qquad r(0) = r_0. \tag{4.36}$$

This ordinary differential equation describes the wealth dynamics in continuous time.

The ordinary differential equation (4.36) can be extended to an open neighborhood of the simplex $\Delta_I$. Namely, since the linear mappings $r \mapsto \sum_{j=1}^{I} r_j \lambda_{j,k}$ in the denominator of (4.34) are continuous on $\mathbb{R}^I$ and strictly positive on $\Delta_I$, $N$ and $\psi$ in (4.34) and (4.35) are defined on an open neighborhood $D$ of $\Delta_I$. Then, for given initial value $r_0 \in D$, the solution of (4.36) exists for all times $t$ smaller than some maximal $t^+(r_0) > 0$ and larger than some minimal $t^-(r_0) < 0$.

We associate a flow

$$\phi : \Gamma \to D, \ (t, r_0) \mapsto \phi_t(r_0) \tag{4.37}$$

with the ordinary differential equation (4.36), where $\phi_t(r_0)$ is the value of the solution of (4.36) at time $t$ when the initial value is $r_0$. Its domain $\Gamma \subseteq \mathbb{R} \times D$ is given by

$$\Gamma = \bigcup_{r \in D} (t^-(r), t^+(r)) \times \{r\} \, .$$

A flow satisfies the following four properties: (1) $\Gamma$ is open in $\mathbb{R}_+ \times D$. (2) $\phi : \Gamma \to D$ is continuous. (3) $\phi_0 = id_D$. (4) For initial value $r \in D$ and times $s \in (t^-(r), t^+(r))$, $t \in (t^-(\phi_s(r)), \phi_s(t^+(r)))$, it holds that $t^-(r) < s + t < t^+(r)$ and $\phi_t(\phi_s(r)) = \phi_{s+t}(r)$.

We need some concepts from convex geometry. The *relative interior* of a convex set $C$ will be denoted by ri $(C)$, i.e.

$$\text{ri } (C) = \{c \in C : \ \exists \epsilon > 0 \ \forall y \in C \quad \forall |\delta| < \epsilon \ \ c + \delta(y - c) \in C\} \, .$$

The *relative boundary* of a convex set $C$ is defined by $\partial^*(C) := \bar{C} \backslash \text{ri } (C)$.

In contrast to the standard topological concept of open sets, the set ri $(C)$ is never empty, whenever the convex set $C$ is not empty. For instance, the set $C = \{x\}$ has relative interior ri $(C) = \{x\}$.

***Definition 4.4.2.*** *A set $M \subseteq D$ is called* invariant*, if $\phi_t(r) \in M$ for all $r \in M$ and $t \in (t^-(r), t^+(r))$. $M$ is called* positively invariant*, if $\phi_t(r) \in M$ for all $r \in M$ and $t \in [0, t^+(r))$.*

It is not difficult to show that $\Delta_I$ is invariant, cf. Amann (1983), Corollary 16.10. This has implications for the domain $\Gamma$ of the flow. Since $\Delta_I$ is compact and invariant, the solution of the differential equation (4.36) exists for all times $t \in \mathbb{R}$, if the initial value $r_0 \in \Delta_I$ ([Amann (1983)], Remark (17.3)). We obtain that

$$\Gamma \; = \; \Big(\mathbb{R} \times \Delta_I\Big) \; \cup \; \left(\bigcup_{r \in D \backslash \Delta_I} (t^-(r), t^+(r)) \times \{r\}\right) . \tag{4.38}$$

Besides the simplex $\Delta_I$ also the sets $\partial^*(\Delta_I)$ and ri $(\Delta_I)$ are invariant; this is implied by standard arguments, cf. Amann (1983), Corollary 16.10. Moreover, the vertices $e_i$ of the simplex $\Delta_I$ are fixed points of the flow. Here, $e_i$ denotes the $i$th unit vector in $\mathbb{R}^I$.

Finally, we define for $J \subseteq I$ the subsimplices

$$\Delta_J := \left\{ \sum_{i \in J} r_i e_i : \ r \in \mathbb{R}_+, \quad \sum_{j \in J} r_j = 1 \right\} .$$

For $J \subseteq I$, $\Delta_J \subseteq \Delta_I$ is invariant. In economic terms, the restriction to a simplex $\Delta_J$, $J \subseteq I$, $J \neq I$ corresponds to a smaller economy where only agents from set $J$ are present. If the initial value is an element of the boundary, i.e. $r \in \partial^*(\Delta_I)$, the wealth dynamics is effectively of lower dimension. Hence, we need to analyze the dynamics for initial values $r \in ri(\Delta_I)$.

### 4.4.2 A Lyapunov Function and LaSalle's Criterion

We will now characterize the asymptotic behavior of the semiflow of the wealth dynamics. For this purpose, we will investigate a Lyapunov function of the flow. Lyapunov functions are defined in terms of derivatives along the orbit of the flow inside a given set $M$, cf. Amann (1983). We do not need this definition in full generality. Instead we will work with the following sufficient criterion that characterizes Lyapunov functions on an open neighborhood of $M$ by their gradient.

***Lemma 4.4.3.*** *Let $M \subseteq D$. A differentiable function $\Phi : U \to \mathbb{R}$, defined on some open neighborhood $U$ of $M$, is a Lyapunov function on $M$ of the semiflow $\phi$ associated with $\psi$ if*

$$\dot{\Phi}(r) := \nabla_r \Phi(r) \psi(r) \leq 0 \qquad \forall r \in M.$$

$\Phi$ is non-increasing along trajectories $\phi_t(r_0)$ for $r_0 \in M$. We recall the following corollary of the invariance principle of LaSalle.

***Corollary 4.4.4.*** *Let $M \subseteq D$ be closed and positively invariant for the semiflow $\phi$. Assume that $\Phi$ is a Lyapunov function on $M$. Let $M_\Phi$ be the largest invariant subset of*

$$\left\{ r \in M : \dot{\Phi}(r) = 0 \right\}.$$

*Then, $M_\phi$ attracts all points of $M$, i.e. for all $r \in M$ we have*

$$\lim_{t \to t^+(r)} \text{dist}\left(\phi_t(r), M_\Phi\right) \ = \ 0 \, .$$

We will next use Corollary 4.4.4 to characterize the minimal attractor of $\Delta_I$. It describes the long-run wealth distribution in the economy, if initially no more than $I$ agents are present. A more detailed analysis allows us to determine the minimal attractor of $ri(\Delta_I)$. This second attractor captures the long-run wealth distribution in the economy, if initially the wealth of *all* $I$ investors is positive, i.e. if initially (and thus for every finite time) exactly $I$ agents are present.

A Lyapunov function $\Phi$ for the flow that describes the wealth dynamics is given in the following lemma.

***Lemma 4.4.5.*** *Suppose that Assumptions 4.2.1 and 4.4.1 are satisfied. The function $\Phi : D \to \mathbb{R}$, defined as*

$$\Phi(r) := -\sum_{k=1}^{K} \pi_k \log(\sum_{j=1}^{I} \lambda_{j,k} r_j) + \sum_{k=1}^{K} \sum_{j=1}^{I} \lambda_{j,k} r_j \ , \tag{4.39}$$

*is a Lyapunov function for the flow $\phi$ on $\Delta_I$ and satisfies on $D$ the equation $N = -\nabla_r \Phi$. The Lyapunov function $\Phi$ is convex on $\Delta_I$.*

*Proof.* The equation $N = -\nabla\Phi$ is easily verified. In particular, we obtain on the simplex $\Delta_I$ the inequality

$$\dot{\Phi}(r) = \nabla_r \Phi(r)\psi(r) = -\sum_{i=1}^{I} r_i N_i^2(r) \leq 0.$$

The convexity of $\Phi$ follows from the concavity of the logarithm. □

The next corollary completely characterizes the long-run wealth distributions in an economy with no more than $I$ agents.

***Corollary 4.4.6.*** *Suppose that Assumptions 4.2.1 and 4.4.1 are satisfied. The minimal attractor of $\Delta_I$ for the flow $\phi$ is given by*

$$\mathcal{A} := \left\{ r \in \Delta_I : \sum_{i=1}^{I} r_i N_i^2(r) = 0 \right\}.$$

*$\mathcal{A}$ is a set of fix points. In particular, for all $r \in \Delta_I$ the $\omega$-limit set $\omega(r)$ is included in $\mathcal{A}$.*

*Proof.* We denote the minimal attractor of $\Delta_I$ by $\tilde{\mathcal{B}}$. The inclusion $\tilde{\mathcal{B}} \subseteq \mathcal{A}$ is implied by LaSalle's Corollary 4.4.4, since $\dot{\Phi}(r) = -\sum r_i N_i^2(r)$. Conversely, the condition $\sum r_i N_i^2(r) = 0$ implies

$$\forall i = 1, 2, \ldots, I : \quad r_i = 0 \ \vee \ N_i(r) = 0.$$

Thus, $\psi(r) = 0$ for $r \in \mathcal{A}$. $\mathcal{A}$ is therefore a set of fix points for the flow $\phi$, hence $\mathcal{A} \subseteq \tilde{\mathcal{B}}$. □

### 4.4.3 The Global Attractor

We will now investigate the asymptotic properties of the solution of the ordinary differential equation (4.36) for initial values $r(0) = r_0 \in ri(\Delta_I)$. Recall that the differential equation (4.36) describes the dynamics of investors' wealth. We are interested in the smallest closed set $\mathcal{B}$ attracting all points $r \in \mathrm{ri}\,(\Delta_I)$ which is given by

$$\mathcal{B} = \overline{\bigcup_{r \in \mathrm{ri}\,(\Delta_I)} \omega(r)}.$$

$\mathcal{B}$ characterizes the long-run wealth distributions in an economy with $I$ agents, if initial wealth of *all* investors is positive. The analysis thus refines Corollary 4.4.6 in which we determined the long-run wealth asymptotics in an economy with *no more* than $I$ agents.

The minimal attractor $\mathcal{B}$ of the relative interior $ri(\Delta_I)$, the attractor $\mathcal{A}$ of the whole simplex $\Delta_I$, and the minima of the Lyapunov function $\Phi$ are closely related. We denote the set of global minima of the Lyapunov function $\Phi$ on $\Delta_I$ by $\mathcal{A}_{\min}$. Since $\Phi$ is a convex function, global and local minima coincide.

***Remark 4.4.7.*** *Elementary relations between the attractors of the simplex and its relative interior and the minima of the Lyapunov function are described in Proposition 4.6.3. In particular, the following holds.*

- $\mathcal{A}_{\min}$ *is a non empty, closed, convex set of fixed points for* $\Phi$.
- *Both* $\mathcal{B}$ *and* $\mathcal{A}_{\min}$ *are subsets of* $\mathcal{A}$.
- *Finally,* $\mathcal{A}_{\min}$ *is a subset of* $\mathcal{B}$*, if* $\mathcal{A}_{\min}$ *contains points of the relative interior* $ri\,(\Delta_I)$.

In certain cases the minimal attractor $\mathcal{B}$ of the relative interior of the simplex can completely be characterized by the minima of the Lyapunov function. In this case, these minima determine the long-run wealth distributions of the dynamics. The next theorem provides conditions.

***Theorem 4.4.8.*** *Suppose that Assumptions 4.2.1 and 4.4.1 are satisfied.*

*Assume that one of the following two conditions is satisfied.*

*(1)* $\Phi$ *is strictly convex on the boundary* $\partial^*(\Delta_I)$*, that is* $\Phi : D \to \mathbb{R}$ *is strictly convex for all convex subsets of the boundary* $\partial^*(\Delta_I)$.

*(2)* $\Phi(e_i) = \min_{g \in \partial^*(\Delta_I)} \Phi(g)$ *for some* $i \in I$.

*Then* $\mathcal{B} \subseteq \mathcal{A}_{\min}$*. If additionally* $\mathcal{A}_{\min}$ *contains points of the relative interior of* $\Delta_I$*, then* $\mathcal{B} = \mathcal{A}_{\min}$.

*Proof.* The theorem is a consequence of Lemma 4.6.5 which is proven in the appendix. First assume that (1) holds. Suppose that $\mathcal{B} \backslash \mathcal{A}_{\min} \neq \emptyset$. By Lemma 4.6.5 there exists a connected set $\mathcal{C} \subseteq \partial^*(\Delta_I)$ satisfying the properties $(C_1)$ and $(C_2)$ and $(C_3)$. The properties $(C_2)$ and $(C_3)$ imply that $\mathcal{C}$ contains at least two points.

Define $M_n = \bigcup_{\substack{J \subseteq I \\ |J| \leq n}} \Delta_J$. Note that $\mathcal{C} \subseteq \partial^*(\Delta_I) = M_{I-1}$. Take the minimal $n$ such that $\mathcal{C} \subseteq M_n$. By minimality of $n$ we find $J \subseteq I$ such that $|J| = n$ and $\mathcal{C} \cap \mathrm{ri}\,(\Delta_J) \neq \emptyset$. Because all points $\mathcal{C} \cap \mathrm{ri}\,(\Delta_J)$ are minima of $\Phi$ on $\Delta_J$ and $\Phi$ is strictly convex on $\Delta_J$, we obtain $|\mathcal{C} \cap \mathrm{ri}\,(\Delta_J)| = 1$, a contradiction, since $|\mathcal{C}| \geq 2$ and $\mathcal{C}$ connected.

Now assume that (2) holds. Again by the subgradient inequality we obtain for all $c \in \mathcal{A} \setminus \mathcal{A}_{\min}$ and for all $i \in I$,

$$N_i(c) = N_i(c) - \sum_{j=1}^{I} c_j N_j(c) = \nabla\Phi(c)(e_i - c) \geq \Phi(c) - \Phi(e_i) > 0.$$

Hence, a set $\mathcal{C}$ as stated in Lemma 4.6.5 satisfying the properties $(C_2)$ and $(C_3)$ simultanously cannot exist.

If additionally $\mathcal{A}_{\min}$ contains points of the relative interior of $\Delta_I$, then the equality $\mathcal{B} = \mathcal{A}_{\min}$ is implied by Proposition 4.6.3(3). $\square$

The next proposition further investigates condition (1) of the preceeding theorem. For this purpose, we define a function

$$\tilde{\Phi} : \begin{cases} (\mathbb{R}_+ \setminus \{0\})^K & \to \quad \mathbb{R} \\ \qquad x & \mapsto \quad -\sum_{k=1}^K \pi_k \log(x_k) + \sum_{k=1}^K x_k. \end{cases} \tag{4.40}$$

The Lyapunov function $\Phi$ can be recovered from $\tilde{\Phi}$ by

$$\Phi(r) = \tilde{\Phi}\left( \left( \sum_{i=1}^{I} r_i \lambda_{i,k} \right)_k \right) \tag{4.41}$$

Recall that by (4.7) the argument of $\tilde{\Phi}$ equals the real price vector $(q_k)_{k=1,2,\dots,K}$ of the assets. The minimization of the Lyapunov function $\Phi$ on the space of wealth distributions consists thus of the two steps: minimize firstly the associated Lyapunov function $\tilde{\Phi}$ on the price space, and find secondly the wealth distributions that support this price vector given the fixed strategy profile.

Next, define the matrix

$$M^{(i)} = (\lambda_j - \lambda_i)_{j \in I \setminus \{i\}} \in \mathbb{R}^{K, I-1} \qquad (i = 1, 2, \dots, I). \tag{4.42}$$

$M^{(i)}$ defines a linear mapping from $\mathbb{R}^{I-1}$ to $\mathbb{R}^K$, and we denote its nullspace by $\ker M^{(i)}$ and its rank by $\mathrm{rg}\ M^{(i)}$. The rank $\mathrm{rg}\ M^{(i)}$ does not depend on the choice of $i$.

***Proposition 4.4.9.*** *Let $d = I - 1 - rg\ M^{(i)} = \dim(\ker M^{(i)})$.*

*(1) If $d = 0$, then $\Phi$ is strictly convex. In particular, $\Phi$ is strictly convex on the boundary $\partial^*(\Delta_I)$.*

*(2) If $d \geq 2$, then $\Phi$ is not strictly convex on the boundary $\partial^*(\Delta_I)$.*

*(3) If $d = 1$, then $\Phi$ is generically strictly convex on the boundary $\partial^*(\Delta_I)$. To be more precise, set $G := \{0, e_1, e_2, \ldots, e_I\} \setminus \{e_i\}$. If $d = 1$, then $\Phi$ is strictly convex on the boundary, if and only if for all $v \in G$,*

$$\ker M^{(i)} \cap span\{u - \bar{u} : \ \ u, \bar{u} \in G \setminus \{v\}\} = \{0\}.$$

*Proof.* W.l.o.g. assume that $i = I$, and set $M := M^{(I)}$.
*Ad (1).* By (4.41), $\Phi(r) = \tilde{\Phi}(L(r))$ with

$$L(r) = M \cdot \begin{pmatrix} r_1 \\ \vdots \\ r_{I-1} \end{pmatrix} + \lambda_I. \tag{4.43}$$

Let $r, r' \in \Delta_I$, $r \neq r'$, $\alpha \in (0,1)$. If $d = 0$, $L$ is injective on $\Delta_I$, thus

$$\begin{aligned} & \Phi(\alpha r + (1-\alpha) r') \ = \ \tilde{\Phi}(\alpha L(r) + (1-\alpha) L(r')) \\ > \quad & \alpha \tilde{\Phi}(L(r)) + (1-\alpha)\tilde{\Phi}(L(r')) \ = \ \alpha \Phi(r) + (1-\alpha)\Phi(r'). \end{aligned}$$

Thus, $\Phi$ is strictly convex.
*Ad (2).* The mapping $L' : \mathbb{R}^I \to \mathbb{R}^K$, $r \mapsto L(r) - \lambda_I$ is linear with $\dim(\ker L') \geq 3$. Let $J = I \setminus \{I\}$, $r \in ri(\Delta_J)$, $N_J = \text{span}\{(e_j - r) : \ \ j \in I \setminus \{I\}\}$. Then $\dim N_J = I - 2$. Thus,

$$\dim(N_J \cap \ker L') \geq \dim N_J + \dim(\ker L') - I = 1.$$

Let $v \in (N_J \cap \ker L') \setminus \{0\}$. Then for $\delta > 0$ sufficiently small, $\epsilon \in [0, \delta]$, we obtain that $r + \epsilon v \in \Delta_J$. Moreover, $L(r + \epsilon v) = L(r)$. This implies $\Phi(r + \epsilon v) = \Phi(v)$. Hence, $\Phi$ is not strictly convex on $\Delta_J$.
*Ad (3).* First, let $\Phi$ be strictly convex on the boundary. Suppose there exists $v \in G$ such that the subspace

$$Q_v = \ker M \cap \text{span}\{u - \bar{u} : \ u, \bar{u} \in G \setminus \{v\}\}$$

is non trivial. Then, we can find $w \in Q_v \setminus \{0\}$.

If $v \neq 0$, then $v = e_j$ for some $j \in I$. Otherwise, let $j = I$. Let $J := I \setminus \{j\}$, $r \in ri(\Delta_J)$. Define

$$\bar{w} = \begin{pmatrix} w \\ -\sum_{k=1}^{I-1} w_k \end{pmatrix}.$$

It is easy to see that $r + \epsilon\bar{w} \in \Delta_J$, if $|\epsilon|$ is sufficiently small, say $|\epsilon| < \delta$ for some $\delta > 0$. We obtain that $L(r + \epsilon\bar{w}) = L(r)$. This implies that $\Phi(r + \epsilon\bar{w}) = \Phi(r)$. Hence, $\Phi$ is not strictly convex on $\Delta_J$.

The converse can be verified as follows. We set $J = I \setminus \{e_j\}$, $j \in I$. Let $r, r' \in \Delta_J$. Suppose that $L(r) = L(r')$. Then

$$w := \begin{pmatrix} r_1 - r'_1 \\ \vdots \\ r_{I-1} - r'_{I-1} \end{pmatrix} \in \ker M.$$

Let $v = e_j$, if $j \neq I$; otherwise, $v = 0$. Then $w \in \operatorname{span}\{u - \bar{u} : \ u, \bar{u} \in G \setminus \{v\}\}$, hence $w = 0$. This implies that $L$ is injective on $\Delta_J$. Analogous to part (1), it follows that $\Phi$ is strictly convex on $\Delta_J$. □

The analysis of the attractor $\mathcal{B}$ of the relative interior of the simplex $\Delta_I$ is more complicated, if $\mathcal{B} \nsubseteq \mathcal{A}_{\min}$. In this case, observe that the long-run wealth distributions $\mathcal{B}$ can be decomposed as

$$\mathcal{B} \ = \ (\mathcal{B} \cap \mathcal{A}_{\min}) \ \cup \ (\mathcal{B} \setminus \mathcal{A}_{\min})\,.$$

Apart from the minima of the Lyapunov function, we thus need to investigate the set $\mathcal{B} \setminus \mathcal{A}_{\min}$.

***Proposition 4.4.10.*** *$\mathcal{B} \setminus \mathcal{A}_{\min}$ is a subset of the boundary $\partial^*(\Delta_I)$. Any $r^* \in \mathcal{B} \setminus \mathcal{A}_{\min}$ is contained in the relative interior $ri\,(\Delta_J)$ of some subsimplex for some $J \subseteq I$, $J \neq I$. $r^*$ minimizes the Lyapunov function $\Phi$ on $ri\,(\Delta_J)$.*

*Proof.* Since $\mathcal{B} \subseteq \mathcal{A}$, we get that $\mathcal{B} \setminus \mathcal{A}_{\min} \subseteq \mathcal{A} \setminus \mathcal{A}_{\min}$. Since ri $(\Delta_I) \cap \mathcal{A} \subseteq \mathcal{A}_{\min}$ by Proposition 4.6.3(1), we obtain that $\mathcal{B} \setminus \mathcal{A}_{\min} \subseteq \partial^*(\Delta_I)$. Since $\mathcal{B} \subseteq \mathcal{A}$, $\mathcal{B} \setminus \mathcal{A}_{\min} \subseteq \mathcal{A} \cap \partial^*(\Delta_I)$. Thus, the properties of $r^*$ are implied by Proposition 4.6.4. □

According to the last proposition the elements of $\mathcal{B} \setminus \mathcal{A}_{\min}$ are included in the relative interior of some subsimplex. The following proposition further investigates the boundary of $\Delta_I$ and provides conditions when boundary points cannot be included in $\mathcal{B} \setminus \mathcal{A}_{\min}$. To simplify the notation, we assume that $r^* \in$ ri $(\Delta_J)$ for some $J$ which includes index $I$. Otherwise, we can relabel the unit vectors. We set $M := M^{(I)}$ as defined in (4.42).

***Proposition 4.4.11.*** *Suppose that*

$$\mathit{span}\ \left((e_j)_{j\in J\setminus\{I\}}, \ker M\right) = \mathbb{R}^{I-1}. \tag{4.44}$$

*Then the following holds.*

*(1) $r^*$ is not included in $\mathcal{B}\setminus\mathcal{A}_{\min}$.*

*(2) If $r^*$ is a minimizer of the Lyapunov function on $ri\,(\Delta_J)$, then $r^*\in\mathcal{A}_{\min}$.*

*Proof.* We prove the second part first. The first part is then an elementary consequence. *Ad (2).* $\Phi(r) = \tilde{\Phi}(L(r))$ with $L$ defined according to equation (4.43). We can find $v\in\ker M$ such that

$$\sum_{j\in J\setminus\{I\}} r_j^*\cdot e_j + v \in ri(conv\{0, e_1, \dots, e_{I-1}\})).$$

Let $u = r^* + \begin{pmatrix} v \\ -\sum_{j=1}^{I-1} v_j \end{pmatrix} \in ri(\Delta_I)$. Then $\Phi(u) = \Phi(r^*)$.

Let $w\in\mathbb{R}^I$ such that $\sum_{j=1}^{I} w_j = 0$. Then there exist $w^1\in \mathrm{span}\{(e_j)_{j\in J\setminus\{I\}}\}\subseteq\mathbb{R}^{I-1}$, $w^2\in\ker M\subseteq\mathbb{R}^{I-1}$ such that

$$\begin{pmatrix} w_1 \\ \vdots \\ w_{I-1} \end{pmatrix} = w^1 + w^2.$$

For $|\epsilon|$ sufficiently small, we obtain that $u + \epsilon\begin{pmatrix} w \\ -\sum_{j=1}^{I-1} w_j \end{pmatrix} \in \mathrm{ri}\,(\Delta_I)$ and that

$$\begin{aligned} \Phi(u+\epsilon w) &= \tilde{\Phi}\left(L\left(\sum_{j=1}^{I-1}(r^* + v + \epsilon w^1 + \epsilon w^2)_j\cdot e_j\right)\right) \\ &= \Phi\left(r^* + \epsilon\begin{pmatrix} w^1 \\ -\sum_{j=1}^{I-1} w_j^1 \end{pmatrix}\right) \ \geq\ \Phi(r^*)\ =\ \Phi(u). \end{aligned}$$

The inequality follows from the fact that $r^* + \begin{pmatrix} w^1 \\ -\sum_{j=1}^{I-1} w_j^1 \end{pmatrix} \in \Delta_J$.

We obtain that $u$ is a local minum of $\Phi$ on $\Delta_I$. Hence, $u$ is a global minimum. From this follows that $r^*\in\mathrm{argmin}_{r\in\Delta_I}\Phi(r)$.
*Ad (1).* Suppose $r^*\in\mathcal{B}\setminus\mathcal{A}_{\min}$. Then $r^*\in\mathrm{argmin}_{r\in ri(\Delta_J)}\Phi(r)$ for some $J\subseteq I$, $J\neq I$ by Proposition 4.4.10. Then part (2) implies that $r^*\in\mathcal{A}$, a contradiction. □

***Remark 4.4.12.*** *Proposition 4.4.11 provides essentially a condition in terms of the dimension of the kernel of the matrix $M$. If $\dim(\ker M)$ is large enough, then (4.44) is generically satisfied.*

### 4.4.4 The Minima of the Lyapunov Function

The Lyapunov function $\Phi$ is the composition of a strictly convex function $\tilde{\Phi}$ and a linear mapping according to (4.41). We investigate the minima $\mathcal{A}_{\min}$ of $\Phi$ in two steps. First we determine the minima of $\tilde{\Phi}$. Then, we investigate the implications for the minima of $\Phi$.

***Lemma 4.4.13.*** *Suppose that Assumptions 4.2.1 and 4.4.1 are satisfied.*

*(1) $\pi$ is the unconstraint absolute minimizer of $\tilde{\Phi}$.*

*(2) We denote by $\Lambda \subseteq \Delta_K$ the convex hull of the trading strategies $\lambda_1, \ldots, \lambda_I$. Then there exists a unique $x^*$ such that*

$$\tilde{\Phi}(x^*) = \inf_{x \in \Lambda} \tilde{\Phi}(x). \tag{4.45}$$

*The minimizer $x^*$ depends on both the real dividends $\pi$ and the polyhedral set $\Lambda$. $x^* = \pi$, if and only if $\pi \in \Lambda$.*

*Proof.* Part (1) is implied by the strict convexity of $\tilde{\Phi}$ and the first order conditions. We only need to show that (2) holds. The set $\Lambda \subseteq \Delta_K$ is a compact set included in ri $(\Delta_K)$. Since $\tilde{\Phi}$ restricted to $\Lambda$ is continuous, there exists a global minimum attained at some point $x^* \in \Lambda$. Assumption (4.4.1) ensures that $\tilde{\Phi}$ is strictly convex. Moreover, $\Lambda$ is convex. This implies the uniqueness of $x^*$. Finally, if $\pi \in \Lambda$, then clearly $x^* = \pi$. Conversely, if $\pi \notin \Lambda$, then $\pi \neq x^* \in \Lambda$. $\square$

Given the minimizer $x^*$ of $\tilde{\Phi}$ on $\Lambda$, the set of minima $\mathcal{A}_{\min}$ of the Lyapunov function $\Phi$ on the simplex $\Delta_I$ is essentially determined by the solution of a linear equation. $\mathcal{A}_{\min}$ is a polyhedral set, that is, the convex hull of finitely many points.

***Lemma 4.4.14.*** *$\mathcal{A}_{\min}$ is a non empty polyhedral set and can be represented by*

$$\mathcal{A}_{\min} = \left\{ r \in \Delta_I : \sum_{i=1}^{I} r_i \lambda_{i,k} = x_k^* \quad \textit{for all } 1 \leq k \leq K \right\}. \tag{4.46}$$

*Proof.* $\mathcal{A}_{\min}$ is non empty by Proposition 4.6.3. Representation (4.46) is an immeditate consequence of (4.41) and (4.45), as the following calculation shows:

$$\min_{r \in \Delta_I} \Phi(r) = \min_{r \in \Delta_I} \tilde{\Phi}\left( \left( \sum_{j=1}^{I} r_j \lambda_{j,k} \right)_{k=1,\ldots,K} \right) = \min_{x \in \Lambda} \tilde{\Phi}(x) = \tilde{\Phi}(x^*). \tag{4.47}$$

The solution of the linear system in $\mathbb{R}^I$, given by $\sum_{i=1}^{I} r_i \lambda_i = x^*$ with $r$ unknown, is an affine subspace of $\mathbb{R}^I$. This implies that $\mathcal{A}_{\min}$ is the intersection of a simplex and an affine subspace, hence polyhedral. $\square$

The next proposition formulates a necessary and sufficient criterion that the minimizer $x^*$ equals a trading strategy $\lambda_i$ for some $i \in I$.

***Proposition 4.4.15.*** *Let $i \in I$. Then the following conditions are equivalent.*

(i) $\lambda_i = x^*$.

(ii) $\nabla\tilde{\Phi}(\lambda_i) \cdot (\lambda_j - \lambda_i) \geq 0$ *for all* $j \in I \setminus \{i\}$.

(iii) $\sum_{k=1}^K \pi_k \frac{\lambda_{j,k}}{\lambda_{i,k}} \leq 1$ *for all* $j \in I \setminus \{i\}$.

*Proof.*
*(i) ⇒ (ii):* Suppose not. Then there exists $j \in I \setminus \{i\}$ such that

$$\nabla\tilde{\Phi}(\lambda_i) \cdot (\lambda_j - \lambda_i) < 0.$$

Define for $\alpha \in [0,1]$ the vector $x(\alpha) := \alpha\lambda_j + (1-\alpha)\lambda_i \in \Lambda$. Then

$$\frac{d}{d\alpha}\tilde{\Phi}(x(\alpha))_{|\alpha=0} = \nabla\tilde{\Phi}(\lambda_i) \cdot (\lambda_j - \lambda_i) < 0.$$

Since $\alpha \mapsto \frac{d}{d\alpha}\tilde{\phi}(x(\alpha))$ is continuous, we can find $1 \geq \epsilon > 0$ such that $\alpha \mapsto \tilde{\Phi}(x(\alpha))$ is strictly decreasing on $[0,\epsilon]$. Thus, $\tilde{\Phi}(x(\epsilon)) < \tilde{\Phi}(\lambda_i)$. This implies $\lambda_i \neq x^*$, a contradiction.
*(ii) ⇒ (i):* Let $x = \sum_{j=1}^I r_j\lambda_j$ with $r_j \geq 0$ $(j \in I)$, $\sum_{j=1}^I r_j = 1$. Then

$$\nabla\tilde{\Phi}(\lambda_i) \cdot (x - \lambda_i) = \sum_{j=1}^I r_j \nabla\tilde{\Phi}(\lambda_i) \cdot (\lambda_j - \lambda_i) \geq 0.$$

Thus, by the subgradient inequality for convex functions

$$\tilde{\Phi}(x) \geq \tilde{\Phi}(\lambda_i) + \nabla\tilde{\Phi}(\lambda_i) \cdot (x - \lambda_i) \geq \tilde{\Phi}(\lambda_i).$$

Since the minimum of $\tilde{\Phi}$ is unique, we obtain $\lambda_i = x^*$.
*(ii) ⇔ (iii):*

$$\begin{aligned}\nabla\tilde{\Phi}(\lambda_i) \cdot (\lambda_j - \lambda_i) \quad &= \quad \sum_{k=1}^K \left(-\frac{\pi_k}{\lambda_{i,k}} + 1\right) \cdot (\lambda_{j,k} - \lambda_{i,k}) \\ = \quad -\sum_{k=1}^K \pi_k \cdot \frac{\lambda_{j,k}}{\lambda_{i,k}} + \sum_{k=1}^K \pi_k \quad &= \quad 1 - \sum_{k=1}^K \pi_k \frac{\lambda_{j,k}}{\lambda_{i,k}}.\end{aligned}$$

This clearly implies the equivalence of (ii) and (iii). □

***Corollary 4.4.16.*** *Assume that one and thus all of the equivalent conditions in Proposition 4.4.15 hold. Then*

$$\mathcal{B} \subseteq \mathcal{A}_{\min} = \left\{ r \in \Delta_I : \sum_{j=1}^{I} r_j \lambda_j = \lambda_i \right\}. \tag{4.48}$$

*If additionally $\mathcal{A}_{\min}$ contains a point in the relative interior $ri\,(\Delta_I)$, then $\mathcal{B} = \mathcal{A}_{\min}$ in (4.48).*

*Proof.* $\mathcal{A}_{\min} = \left\{ r \in \Delta_I : \sum_{j=1}^{I} r_j \lambda_j = \lambda_i \right\}$ by (4.46). Then $e_i \in \mathcal{A}_{\min}$, hence $\mathcal{B} \subseteq \mathcal{A}_{\min}$ by Theorem 4.4.8(2). The last claim follows from Proposition 4.6.3. □

***Corollary 4.4.17.*** *Assume that one and thus all of the equivalent conditions in Proposition 4.4.15 hold. If $\lambda_i$ is an extremal point $\Lambda$, then*

$$\mathcal{B} = \mathcal{A}_{\min} = \{e_i\}.$$

*Proof.* If $\lambda_i$ is an extremal point of the polyhedron $\Lambda$, then $\mathcal{A}_{\min} = \{e_i\}$. Since $\emptyset \neq \mathcal{B} \subseteq \mathcal{A}_{\min}$, we obtain $\mathcal{B} = \mathcal{A}_{\min}$. □

In general, the minimizer $x^*$ of $\tilde{\Phi}$ is a convex combination of the trading strategy $\lambda_j$, $j \in I$. The next proposition characterizes trading strategies which will never contribute to $x^*$.

***Proposition 4.4.18.*** *Assume that for some $i \in I$ the following inequality is satisfied, namely*

$$\sum_{k=1}^{K} \pi_k \frac{\lambda_{i,k}}{x_k^*} \neq 1. \tag{4.49}$$

*If $\sum_{j=1}^{I} r_j \lambda_j = x^*$ or, equivalently, $r \in \mathcal{A}_{\min}$ for some $r \in \Delta_I$, then $r_i = 0$.*

*Proof.* Since $\nabla\tilde{\Phi}(x^*) \cdot (\lambda_i - x^*) = 1 - \sum_{k=1}^{K} \pi_k \cdot \frac{\lambda_{i,k}}{x_k^*}$, we obtain

$$\nabla\tilde{\Phi}(x^*) \cdot (\lambda_i - x^*) \neq 0.$$

Let now $y \in \Lambda$. Assume that

$$\nabla\tilde{\Phi}(x^*) \cdot (y - x^*) < 0.$$

For $\alpha \in [0,1]$ define the vector $x(\alpha) := \alpha y + (1-\alpha)x^* \in \Lambda$. The same arguments as in the part (i)⇒(ii) of the proof of Proposition 4.4.15 show that there exists $0 < \epsilon \leq 1$

such that $\tilde{\Phi}(x(\epsilon)) < \tilde{\Phi}(x^*)$. This implies that $x^* \neq \operatorname{argmin}_{x\in\Lambda}\tilde{\Phi}(x)$, a contradiction. Hence, for $y \in \Lambda$,

$$\nabla\tilde{\Phi}(x^*)\cdot(y - x^*) \;\geq\; 0, \tag{4.50}$$

$$\nabla\tilde{\Phi}(x^*)\cdot(\lambda_i - x^*) \;>\; 0. \tag{4.51}$$

Now, let $r \in \Delta_I$ such that $x^* = \sum_{j=1}^{I} r_j\lambda_j$. Then,

$$0 = \nabla\tilde{\Phi}(x^*)\cdot(x^* - x^*) = \sum_{j=1}^{I} r_j\nabla\tilde{\Phi}(x^*)\cdot(\lambda_j - x^*).$$

Since each summand is non negative by (4.50), we obtain that

$$r_i\nabla\tilde{\Phi}(x^*)\cdot(\lambda_i - x^*) = 0.$$

Finally, (4.51) implies that $r_i = 0$. □

Suppose now that $r_0$ is an initial value of the wealth distribution among investors with asymptotics $\omega(r_0) \subseteq \mathcal{A}_{\min}$. If the condition (4.49) of the preceding proposition is satisfied, then strategy $\lambda_i$ dies out in the long run, that is, $r_i = 0$ for $r \in \omega(r_0)$. Condition (4.49) depends on $\Lambda$: whether a trading strategy dies out or not for initial value $r_0$ with $\omega(r_0) \subseteq \mathcal{A}_{\min}$, is determined by its business environment of competing trading strategies.

### 4.4.5 A Rational Benchmark

The vector $\pi$ is the unconstraint minimizer of the function $\tilde{\Phi}$. This implies that whenever $\pi \in \Lambda$, then

$$\mathcal{A}_{\min} = \left\{ r \in \Delta_I : \sum_{j=1}^{I} r_j\lambda_{j,k} = \pi_k \text{ for all } 1 \leq k \leq K \right\}.$$

By (4.7) the vector $\pi$ equals the price vector $(q_k)_{k=1,2,\ldots,K}$ for any wealth distribution $r \in \mathcal{A}_{\min}$ and the given profile of trading strategies. Under conditions which we already discussed in previous sections the long-run wealth distributions $\mathcal{B}$ are characterized by $\mathcal{A}_{\min}$.

In this section we will compare our results to a rational benchmark of maximizing investors who are price takers in the Walrasian market. In contrast to the evolutionary perspective agents can now choose their trading strategies. It turns out that also in this context the vector $\pi$ plays a special role.

We consider myopic agents who are price takers in a continuous time Walrasian market. The aim of the agents is to maximize the instantaneous gain or growth of their portfolio. By (4.36) and (4.34) the objective function of the agents $i = 1, 2, \ldots, I$ is thus equal to

$$V_i^q : \begin{cases} \Delta_K & \to \quad \bar{\mathbb{R}}_+ \\ \lambda_i & \mapsto \quad \sum_{k=1}^K \frac{\pi_k}{q_k} \lambda_{i,k} - 1 \end{cases}$$

Here, $(q_k)_{k=1,2,\ldots,K}$ equals by (4.7) the real price vector of the assets. In terms of the price, the market clearing conditions can be rewritten as

$$q_k = \sum_{i=1}^{I} \lambda_{i,k} \cdot r_i, \qquad k = 1, 2, \ldots, K. \tag{4.52}$$

Under these conditions we obtain the following result.

***Proposition 4.4.19.*** *The set of Walrasian equlibria in the economy of price taking myopic investors is given by*

$$\mathcal{E} = \left\{ (\lambda_1, \lambda_2, \ldots, \lambda_I) \in (\Delta_K)^I : \sum_{i=1}^{I} \lambda_{i,k} r_i = \pi_k \textit{ for all } 1 \leq k \leq K \right\}.$$

*In equilibrium the price vector $q$ equals $\pi$. Moreover, in equilibrium the wealth vector $(r_i)_{i=1,2,\ldots,I}$ of the investors is constant.*

*Proof.* Clearly, $q_k \neq 0$ for all $k = 1, 2, \ldots, K$. Namely, if $q_k = 0$, the demand for asset $k$ is strictly positive (even infinite). Thus, $q_k \neq 0$ by (4.52), a contradiction. Assume that $\frac{\pi_k}{q_k} > \frac{\pi_l}{q_l}$ for some $l, k = 1, 2, \ldots, K$. Then clearly $\lambda_{j,l} = 0$ for $j = 1, 2, \ldots, I$, since agents are maximizers. Thus, $q_k = 0$ by (4.52), a contradiction. Hence, $\frac{\pi_k}{q_k} = \frac{\pi_l}{q_l}$ $(l, k = 1, 2, \ldots, K)$. Since $\sum_{k=1}^K \pi_k = \sum_{k=1}^K q_k$, this implies that $\pi_k = q_k$ $(k = 1, 2, \ldots, K)$. By (4.52) we obtain that

$$\mathcal{E} \subseteq \left\{ (\lambda_1, \lambda_2, \ldots, \lambda_I) \in (\Delta_K)^I : \sum_{i=1}^{I} \lambda_{i,k} r_i = \pi_k \text{ for all } 1 \leq k \leq K \right\} =: \mathcal{E}'.$$

Clearly, for $(\lambda_1, \lambda_2, \ldots, \lambda_I) \in \mathcal{E}'$ the price vector $q$ equals $\pi$. For $q = \pi$ agents are indifferent between all strategies, thus no profitable deviation exists for any agent. Hence, $\mathcal{E}' \subseteq \mathcal{E}$.

For $(\lambda_1, \lambda_2, \ldots, \lambda_I) \in \mathcal{E}$ we obtain $q = \pi$, thus $V_i^q(\lambda_i) = 0$ for all $i = 1, 2, \ldots, I$. This implies that the wealth vector $(r_i)_{i=1,2,\ldots,I}$ of the investors is constant. □

Finally, observe that the set of Walrasian equilibria for given wealth vector $(r_i)_{i=1,2,\ldots,I}$,

$$\mathcal{E} = \left\{ (\lambda_1, \lambda_2, \ldots, \lambda_I) \in (\Delta_K)^I : \sum_{j=1}^{I} \lambda_{j,k} r_j = \pi_k \text{ for all } 1 \leq k \leq K \right\},$$

and the set of minima of the Lyapunov function $\Phi$,

$$\mathcal{A}_{\min} = \left\{ r \in \Delta_I : \sum_{j=1}^{I} r_j \lambda_{j,k} = \pi_k \text{ for all } 1 \leq k \leq K \right\},$$

for given strategy profile $(\lambda_i)_{i=1,2,\ldots,I}$ with $\pi \in \Lambda$, are dual with respect to each other.

***Remark 4.4.20.*** *Instead of price taking investors who maximize their objective functions $V_i^q$ for given price vector $q$ we could investigate an oligopolistic market game. In this situation the objective function of investors $i = 1, 2, \ldots, I$ equals*

$$U_i(\lambda_1, \lambda_2, \ldots, \lambda_I) = \sum_{k=1}^{K} \frac{\pi_k \lambda_{i,k}}{\sum_{j=1}^{I} r_j \lambda_{j,k}}.$$

*In the Nash equilibrium of the strategic game, each investor $i \in I$ chooses her optimal $\lambda_i$ given the trading strategies of the others. It can be shown that in this strategic situation the unique Nash equilibrium is the strategy profile $(\lambda_1, \lambda_2, \ldots \lambda_I) = (\pi, \pi, \ldots, \pi)$.*

## 4.5 Conclusion

In this chapter, we have discussed a continuous time approximation for the evolutionary stock market model of Blume and Easley (1992). We provided conditions for the convergence of the Euler scheme to a nonlinear integral equation in a random environment. If dividend payments are increments of an excess value process of firms, the analysis reveals that a representation of the value process in terms of a locally finite kernel is useful. In particular, the Euler scheme converges, if the representing kernel dominates the Lebesgue measure, or – alternatively – if the representing densities are smooth enough.

For constant asset return, we investigate the long-run asymptotics of the continuous time wealth process.[2] In this case the integral equation reduces to an autonomous ordinary differential equation. The asymptotic behavior can be characterized by the minima of a Lyapunov function. This relationship was analyzed in detail in Section 4.4.

Finally, we have investigated a rational benchmark. In the context of the evolutionary dynamics, the dividend vector $\pi$ is closely related to the absolute minimizers of the Lyapunov function. For rational investors, this dividend vector determines the set of Walrasian equilibria.

[2]The long-run asymptotics for stochastic dividend processes will be analyzed in future work, cf. Buchmann and Weber (2004b).

## 4.6 Appendix

### 4.6.1 The Dividend Process

For technical reasons, we need the following concept of an *exhausting sequence*.

***Definition 4.6.1.*** *Let $P$ be a probability measure and $\mu$ be a locally finite kernel from $\Omega$ to $\mathbb{R}_+$. A sequence $(C_N)_{N\in\mathbb{N}} \subseteq \mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+)$ is called* exhausting *for $P$ and $\mu$, if the following properties are satisfied.*

*(1)* $C_N \in \{F \times [0, \beta) :\ F \in \mathcal{F},\ \beta > 0\}$ *for all $N$.*

*(2)* $P\mu(C_N) < \infty$.

*(3)* $\bigcup_N C_N = \Omega \times \mathbb{R}_+$.

***Lemma 4.6.2.*** *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\mu$ be a locally finite transition kernel from $\Omega$ to $\mathbb{R}^+$. Then there exists an exhausting sequence $(C_N)_{N\in\mathbb{N}}$ for $P$ and $\mu$. Thus, Definition (4.30) defines a unique $\sigma$-finite measure $P\mu$ on the whole $\sigma$-algebra $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+)$.*

*Proof.* For $n, m \in \mathbb{N}$ define $B_{n,m} := \{\omega :\ \mu(\omega, [0, m)) \leq n\} \times [0, m)$. By definition

$$P\mu(B_{n,m}) = \int_{\mu(\omega,[0,m))\leq n} \mu(\omega, [0, m))\, P(d\omega) \leq n\,.$$

Since $\mu(\omega, \cdot)$ is a locally finite measure on $\mathcal{B}(\mathbb{R}_+)$, we find

$$\bigcup_n B_{n,m} = \{\omega :\ \mu(\omega, [0, m)) < \infty\} \times [0, m) = \Omega \times [0, m)\,.$$

Thus, $\bigcup_{n,m} B_{n,m} = \Omega \times \mathbb{R}_+$.

Finally, Caratheodory's extension theorem implies that a $\sigma$-finite measure $P\mu$ on $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+)$ is uniquely specified by Definition 4.30. $\square$

### 4.6.2 Results Related to the Global Attractor

***Proposition 4.6.3.*** *Suppose that Assumptions 4.2.1 and 4.4.1 are satisfied. Then*

$$\overline{ri\,(\Delta_I) \cap \mathcal{A}} \subseteq \mathcal{B} \subseteq \mathcal{A}.$$

*Moreover, $\mathcal{A}_{\min} \subseteq \mathcal{A}$ is a non empty, closed, convex set of fixed points for $\phi$, and the following holds:*

*(1)* $\overline{ri\,(\Delta_I) \cap \mathcal{A}} \subseteq \mathcal{A}_{\min}$.

*(2) The converse inclusion* $\mathcal{A}_{\min} \subseteq \overline{ri\,(\Delta_I) \cap \mathcal{A}}$ *holds, if and only if the set* $\mathcal{A}_{\min} \cap ri\,(\Delta_I)$ *is non empty.*
*In this case,* $\mathcal{A}_{\min} = \overline{ri\,(\Delta_I) \cap \mathcal{A}} = \overline{ri\,(\Delta_I) \cap \mathcal{A}_{\min}}$.

*(3) Moreover, if* $\mathcal{A}_{\min} \cap ri\,(\Delta_I)$ *is non empty, then* $\mathcal{A}_{\min} \subseteq \mathcal{B}$.

*Proof.* Because $\mathcal{A}$ is a set of fixed points, ri $(\Delta_I) \cap \mathcal{A} \subseteq \mathcal{B}$. As $\mathcal{B}$ is closed, the first inclusion is proved. The second inclusion $\mathcal{B} \subseteq \mathcal{A}$ is a consequence of Corollary 4.4.6, since $\mathcal{A}$ is closed.

Since the Lyapunov function $\Phi$ is continuous on $\Delta_I$, $\mathcal{A}_{\min}$ is closed and non empty. The convexity of $\Phi$ implies that $\mathcal{A}_{\min}$ is convex.

We claim that $\mathcal{A}_{\min} \subseteq \mathcal{A}$. Otherwise, by continuity of $N$ there exist an initial value $r_0 \in \mathcal{A}_{\min}$ with a neighborhood $\mathcal{N}(r_0)$ and $\delta > 0$ such that

$$\sum_{i=1}^{I} r_i N_i(r)^2 \; > \; \delta > \; 0$$

for all $r \in \mathcal{N}(r_0) \cap \Delta_I$. As the flow is continuous, we can find $\epsilon > 0$ such that $\phi_t(r_0) \in \mathcal{N}(r_0)$ for all $0 \le t \le \epsilon$. Therefore

$$\begin{aligned} \Phi(\phi_\epsilon(r_0)) &= \Phi(r_0) + \int_0^\epsilon \dot{\Phi}(\phi_s(r_0))\, ds \\ &= \Phi(r_0) - \int_0^\epsilon \sum_{i=1}^{I} \phi_{i,s}(r_0) N_i(\phi_s(r_0))^2\, ds \\ &\le \Phi(r_0) - \delta\,\epsilon \; < \Phi(r_0), \end{aligned}$$

a contradiction. It follows that $\mathcal{A}_{\min} \subseteq \mathcal{A}$. Hence, $\mathcal{A}_{\min}$ is a set of fixed points.
*Ad (1).* W.l.o.g. suppose ri $(\Delta_I) \cap \mathcal{A} \neq \emptyset$. If $r \in$ ri $(\Delta_I) \cap \mathcal{A}$, then $N_i(r) = 0$ for all $i \in I$ by Corollary 4.4.6. Thus, for all $s \in \Delta_I$

$$\Phi(s) = \Phi(s) + \sum_{i=1}^{I} (s_i - r_i) N_i(r) = \Phi(s) - \nabla\Phi(r)(s - r) \ge \Phi(r)$$

by the subgradient inequality for convex functions. Hence $r \in \mathcal{A}_{\min}$. Since $\mathcal{A}_{\min}$ is closed, the claim follows.
*Ad (2).* If $\mathcal{A}_{\min} \subseteq \overline{\mathcal{A} \cap \text{ri}\,(\Delta_I)}$, then $\mathcal{A} \cap \text{ri}\,(\Delta_I) \neq \emptyset$. However, the points in $\mathcal{A} \cap \text{ri}\,(\Delta_I)$ are minima. Hence $\mathcal{A}_{\min} \cap \text{ri}\,(\Delta_I) \neq \emptyset$.

On the other hand, if $\mathcal{A}_{\min} \cap \text{ri}\,(\Delta_I) \neq \emptyset$, let $r \in \mathcal{A}_{\min} \cap \text{ri}\,(\Delta_I)$. If $s \in \mathcal{A}_{\min}$, then $\alpha s + (1-\alpha) r \in \mathcal{A}_{\min} \cap ri(\Delta_I)$ for $\alpha \in (0, 1]$. Hence, $\mathcal{A}_{\min} = \overline{\mathcal{A}_{\min} \cap \text{ri}\,(\Delta_I)}$. Finally, observe that $\mathcal{A}_{\min} \subseteq \mathcal{A}$.
*Ad (3).* The claim is immediate from $\overline{\text{ri}\,(\Delta_I) \cap \mathcal{A}} \subseteq \mathcal{B}$ and part (2). □

***Proposition 4.6.4.*** *Any $r \in \mathcal{A} \cap \partial^*(\Delta_I)$ is contained in the relative interior $ri\,(\Delta_J)$ for some $J \subseteq I$, $J \neq I$. $r$ minimizes the Lyapunov function $\Phi$ on $\Delta_J$.*

*Proof.* $r$ is clearly contained in the relative interior of some subsimplex.

Moreover, we have that $r = \sum_{i \in J} \; r_i e_i$, $r_i > 0$ $(i \in J)$. Since $r \in \mathcal{A}$, we obtain $N_i(c) = 0$ for all $i \in J$. Hence, for all $x = \sum_{i \in J} x_i e_i \in \Delta_J$

$$\Phi(x) = \Phi(x) + \sum_{j \in J} (x_i - r_i) N_i(r) = \Phi(x) - \nabla\Phi(r)(x - r) \geq \Phi(r)$$

by the subgradient inequality for convex functions. □

We need the following technical lemma.

***Lemma 4.6.5.*** *Let Assumptions 4.2.1 and 4.4.1 be satisfied. Assume that $\mathcal{B} \backslash \mathcal{A}_{\min} \neq \emptyset$. There exist $g \in \mathcal{B} \setminus \mathcal{A}_{\min}$ and $r \in ri(\Delta_I)$ with $g \in \omega(r)$. $\mathcal{C} := \omega(r) \subseteq \mathcal{B} \backslash \mathcal{A}_{\min} \cap \partial^*(\Delta_I)$ is a non empty, connected set satisfying the following properties:*

$(C_1)$ $\forall c \in \mathcal{C}\ \forall J \subseteq I \left( c \in ri\,(\Delta_J) \ \Rightarrow\ \Phi(c) = \min_{d \in \Delta_J} \ \Phi(d) \right)$.

$(C_2)$ $\forall c \in \mathcal{C}\ \exists i \in I\ N_i(c) > 0$.

$(C_3)$ $\forall c \in \mathcal{C}\ \forall i \in I \left( N_i(c) > 0 \ \Rightarrow\ \exists d \in \mathcal{C}\ N_i(d) = 0 \right)$.

*Proof.* Let $\mathcal{B} \backslash \mathcal{A}_{\min} \neq \emptyset$. Then there exist $g \in \mathcal{B} \backslash \mathcal{A}_{\min}$ and $r \in \text{ri}\ (\Delta_I)$ such that $g \in \omega(r)$. Suppose not: Then $\Phi(g) = \min_{x \in \Delta_I} \Phi(x)$ for all $r \in \text{ri}\ (\Delta_I)$ and for all $g \in \omega(r)$. Because $\Phi$ is continuous, $\Phi(s) = \min_{x \in \Delta_I} \Phi(x)$ for all $s \in \mathcal{B}$. Thus, $\mathcal{B} \backslash \mathcal{A}_{\min} = \emptyset$, a contradiction.

Define $\mathcal{C} := \omega(r)$. Since $\Delta_I$ is compact and invariant, $\mathcal{C}$ is a non empty connected set contained in $\Delta_I$ ([Amann (1983)], Theorem (17.2)). Moreover, $\mathcal{C} \subseteq \mathcal{B} \backslash \mathcal{A}_{\min}$, since $\Phi$ is constant on $\omega(r)$. The inclusion $\mathcal{C} \subseteq \partial^*(\Delta_I)$ follows from $\mathcal{C} \subseteq \mathcal{A}$ and property $(C_2)$ proven below.

*Ad $(C_1)$.* Property (1) is a direct consequence of Proposition 4.6.4.

*Ad $(C_2)$.* Suppose that there exists $c = \sum_{i \in I} \; c_i e_i \in \mathcal{C}$ such that $N_i(c) \leq 0$ for all $i \in I$. Let $x = \sum_{i \in J} x_i e_i \in \Delta_I$. Because $c \in \mathcal{A}$, we obtain

$$\sum_{i \in I} (x_i - c_i)\ N_i(c) = \sum_{i \in I} x_i\ N_i(c) \leq 0.$$

Hence, with the same argument as in (1)

$$\Phi(x) \geq \Phi(x) + \sum_{i \in I} (x_i - c_i) N_i(c) = \Phi(x) - \nabla\Phi(c)(x - c) \geq \Phi(c).$$

Thus, $c \in \mathcal{A}_{\min}$, a contradiction.

*Ad ($C_3$).* Let $c \in \mathcal{C}$ and $N_i(c) > 0$. Since $c \in \mathcal{A}$, it follows that $c_i = 0$. Since $c \in \omega(r)$, we find a strictly increasing sequence $(t_{2k}) \subseteq \mathbb{R}_+$ such that $\lim_{k\to\infty} t_{2k} = \infty$ and $\lim_{k\to\infty} \phi_{t_{2k}}(r) = c$. Since $0 = c_i = \lim_{k\to\infty} \phi_{i,t_{2k}}(r)$, we may assume that for all $k \in \mathbb{N}_0$

$$\phi_{i,t_{2(k+1)}}(r) \leq \phi_{i,t_{2k}}(r). \tag{4.53}$$

Recall that $D$ is the open extended state space of the flow as defined in (4.37). Let

$$G = \{g \in D : \; N_i(g) > 0\}.$$

Then $G$ is an open neighborhood of $c$, and therefore $\phi_{t_{2k}}(r) \in G$ for $k$ sufficiently large. W.l.o.g. assume that $\phi_{t_{2k}}(r) \in G$ for all $k$. Define the exit time from the set $G$ by

$$t_{2k+1} = \inf\{t \geq t_{2k} : \; N_i(\phi_t(r)) \leq 0.\}$$

Note that for all $s \in [t_{2k}, t_{2k+1})$

$$\dot{\phi}_{i,s}(r) = \phi_{i,s}(r) \; N_i(\phi_s(r)) > 0.$$

Therefore, $[t_{2k}, t_{2k+1}) \ni s \mapsto \phi_{i,s}(r)$ is strictly increasing. Since (4.53) holds, we obtain that $t_{2k+1}$ must be strictly smaller that $t_{2(k+1)}$. Thus, $t_{2k} < t_{2k+1} < t_{2k+2}$. By continuity of $N$ we obtain that $N_i(\phi_{t_{2k+1}}(r)) = 0$ for all $k$. Because $\{g \in \Delta_I : N_i(g) = 0\} := \mathcal{C}'$ is compact, there exists an element $d \in \mathcal{C}'$ such that it is the limit of an appropriate subsequence of $\phi_{t_{2k+1}}(r)$, namely $d = \lim_{k'\to\infty} \phi_{t_{2k'+1}}(r)$, where $k'$ is some sequence of natural numbers converging to infinity. We obtain that $d \in \mathcal{C} = \omega(r)$ with $N_i(d) = 0$. Hence, property ($C_3$) is proven. $\square$

# Bibliography


Algoet, P. H. and Cover, T. M. (1988). Asymptotic optimality and asymptotic equilibrium. *Annals of Pobability*, 16:876–898.

Allen, F. and Gale, D. (2000). Financial contagion. *Journal of Political Economy*, 108:1–33.

Amann, H. (1983). *Gewöhnliche Differentialgleichungen*. Walter de Gruyter, Berlin.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., and Ku, H. (2003). Coherent multiperiod risk adjusted values and Bellman's principle. Working paper, ETH Zürich.

Aurell, E., Baviera, R., Hammarlid, O., Serva, M., and Vulpiani, A. (2000). Growth optimal investment and pricing of derivatives. *Physica A*, 280:505–521.

Blume, L. and Easley, D. (1992). Evolution and market behavior. *Journal of Economic Theory*, 58:9–40.

Browne, S. and Whitt, W. (1996). Portfolio choice and the Bayesian Kelly criterion. *Advances in Applied Probability*, 28:1145–1176.

Buchmann, B. and Weber, S. (2004a). A continuous time approximation of an evolutionary stock market model. Working Paper, Humboldt-Universität zu Berlin and Australian National University.

Buchmann, B. and Weber, S. (2004b). An evolutionary financial market - a Markovian dynamics. Working Paper, Humboldt-Universität zu Berlin and Australian National University (in preparation).

Carlier, G. and Dana, R.-A. (2003). Core of convex distortions of a probability. *Journal of Economic Theory*, 113:199–222.

Cheridito, P., Delbaen, F., and Kupper, M. (2003). Coherent and convex risk meaures for càdlàg processes. Working paper, ETH Zürich.

Cover, T. M. (1991). Universal portfolios. *Mathematical Finance*, 1:1–29.

Csiszar, I. (1975). *I*-divergence geometry of probability distributions and minimization problems. *Annals of Probability*, 3(1).

Davis, M. and Lo, V. (2001). Modeling default correlation in bond portfolios. In Alexander, C., editor, *Mastering Risk Volume 2: Applications*, pages 141–151. Financial Times Prentice Hall.

Delbaen, F. (2000). Coherent risk measures. Lecture Notes, Scuola Normale Superiore di Pisa.

Delbaen, F. (2002). Coherent risk measures on general probability spaces. In Sandmann, K. and Schönbucher, P. J., editors, *Advances in Finance and Stochastics*, pages 1–38. Springer-Verlag Berlin.

Delbaen, F. (2003). The structure of m-stable sets and in particular of the set of risk-neutral measures. Working paper, ETH Zürich.

Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. Springer, New York.

Detlefsen, K. (2003). Bedingte und mehrperiodische Risikomaße. Diplomarbeit, Humboldt-Universität zu Berlin.

Duffie, D., Schroder, M., and Skiadas, C. (1996). Recursive valuation of defaultable securities and the timing of resolution of uncertainty. *Annals of Applied Probability*, 6:1075–1090.

Duffie, D. and Singleton, K. J. (1999). Modeling term structures of defaultable bonds. *Review of Financial Studies*, 12:687–720.

Durrett, R. (1996). *Probability: Theory and Examples*. Duxbury Press, Belmont.

Ellis, R. S. (1988). Large deviation for the empirical measure of a Markov chain with an application to the multivariate empirical measure. *Annals of Probability*, 16:1496–1508.

Epstein, L. G. and Schneider, M. (2003). Recursive multiple-priors. *Journal of Economic Theory*, 113:1–31.

Evstigneev, I. V., Hens, T., and Schenk-Hoppé, K. R. (2002). Market selection of financial trading strategies: Global stability. *Mathematical Finance*, 12(4):329–339.

Fama, E. F. (1965). The behavior of stock market prices. *Journal of Business*, 38:34–105.

Föllmer, H. and Schied, A. (2002a). Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–448.

Föllmer, H. and Schied, A. (2002b). Robust preferences and convex meaures of risk. In Sandmann, K. and Schönbucher, P. J., editors, *Advances in Finance and Stochastics*, pages 39–56. Springer-Verlag Berlin.

Föllmer, H. and Schied, A. (2002c). *Stochastic Finance - An Introduction in Discrete Time*. Walter de Gruyter, Berlin.

Föllmer, H. and Schied, A. (2004). *Stochastic Finance - An Introduction in Discrete Time (2nd edition)*. Walter de Gruyter, Berlin.

Frey, R. and McNeil, A. J. (2001). Modeling dependent defaults. Working Paper, Department of Mathematics, ETH Zürich.

Frey, R. and McNeil, A. J. (2002). VaR and expected shortfall in portfolios of dependent credit risks: Conceptual and practical insights. *Journal of Banking and Finance*, 26:1317–1334.

Friedman, M. (1953). *Essays in Positive Economics*. University of Chicago Press, Chicago.

Frittelli, M. and Rosazza, G. E. (2002). Putting order in risk measures. *Journal of Banking and Finance*, 26(7):1473–1486.

Fu, M. C., Jin, X., and Xiong, X. (2003). Probabilistic error bounds for simulation quantile estimators. *Management Science*, 49(2).

Georgii, H.-O. (1988). *Gibbs Measures and Phase Transitions*. Walter de Gruyter, Berlin.

Giesecke, K. (2001). Correlated default with incomplete information. Working Paper 30, SFB 373, Department of Economics, Humboldt-Universität zu Berlin.

Giesecke, K. and Weber, S. (2004a). Credit contagion and aggregate losses. Working Paper, Cornell University and Humboldt-Universität zu Berlin.

Giesecke, K. and Weber, S. (2004b). Cyclical correlations, credit contagion, and portfolio losses. *Journal of Banking and Finance*. To appear.

Gordy, M. (2001). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation*. To appear.

Hakansson, N. H. (1970). Optimal investment and consumption strategies under risk for a class of utility functions. *Econometrica*, 38:587–607.

Hakansson, N. H. and Ziemba, W. T. (1995). *Capital Growth Theory*, volume 9, chapter 3, pages 65–86. Elsevier Science.

Hens, T. and Schenk-Hoppé, K. R. (2004). Evolution of portfolio rules in incomplete markets. *Journal of Mathematical Economics*. To appear.

Jarrow, R. A., Lando, D., and Turnbull, S. M. (1997). A Markov model of the term structure of credit risk spreads. *Review of Financial Studies*, 10(2):481–523.

Jarrow, R. A. and Turnbull, S. M. (1995). Pricing derivatives on financial securities subject to credit risk. *Journal of Finance*, 50(1):53–86.

Jarrow, R. A. and Yu, F. (2001). Counterparty risk and the pricing of defaultable securities. *Journal of Finance*, 56(5):555–576.

Jaschke, S. and Küchler, U. (2001). Coherent risk measures and good-deal bounds. *Finance and Stochastics*, 5(2):181–200.

Kallenberg, O. (1997). *Foundations of Modern Probability*. Springer, New York.

Karatzas, I. and Shreve, S. E. (1998). *Methods of Mathematical Finance*. Springer-Verlag, New York.

Kiyotaki, N. and Moore, J. (1997). Credit chains. Working Paper, London School of Economics.

Kondo, K. and Hara, T. (1987). Critical exponent of susceptibility for a class of general ferromagnets in $d > 4$ dimensions. *Journal of Mathematical Physics*, 28(5):1206–1208.

Kunze, M. (2003). Verteilungsinvarinate konvexe Risikomaße. Diplomarbeit, Humboldt-Universität zu Berlin.

Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics*, 3:83–95.

Lando, D. (1998). On Cox processes and credit risky securities. *Review of Derivatives Research*, 2:99–120.

Liggett, T. (1985). *Interacting Particle Systems*. Springer-Verlag, Berlin.

Liggett, T. (1999). *Stochastic Interacting Systems: Contact, Voter, and Exclusion Processes*. Springer-Verlag, Berlin.

Lucas, A., Klaassen, P., Spreij, P., and Straetmans, S. (2001). An analytic approach to credit risk of large corporate bond and loan portfolios. *Journal of Banking and Finance*, 25:1635–1664.

Owen, A. B. (1987). Nonparametric conditional estimation. Ph.D. Thesis, Stanford University.

Rachev, S. T. (1991). *Probability Metrics and the Stability of Stochastic Models*. John Wiley & Sons, Chichester, New York.

Riedel, F. (2002). Dynamic coherent risk measures. Working Paper, Stanford University.

Scandolo, G. (2003). Risk measures in a dynamic setting. Ph.D. Thesis, Università degli Studi Milano & Università di Firenze.

Schloegl, L. (2002). The large homogeneous portfolio approximation with the Student-t copula. Working Paper, Lehman Brothers International, London.

Schmeidler, D. (1986). Integral representation without additivity. *Proceedings American Mathematical Society*, 97(2):255–261.

Schönbucher, P. and Schubert, D. (2001). Copula-dependent default risk in intensity models. Working Paper, Universität Bonn.

Stroock, D. W. (1984). *An Introduction to the Theory of Large Deviations*. Springer, Berlin.

Thorp, E. (1971). Portfolio choice and the Kelly criterion. In R. G. Vickson and W. T. Ziemba, editor, *Stochastic Models in Finance*, pages 599–619. Academic Press, New York.

Vasicek, O. (1987). The loan loss distribution. Technical Document, KMV Corporation.

Wang, T. (1996). A characterization of dynamic risk measures. Working paper, U.B.C.

Wang, T. (1999). A class of dynamic risk measures. Working paper, U.B.C.

Winkler, G. (1985). *Choquet order and simplices with applications to probabilistic models*. Springer-Verlag, Berlin.

Zähle, I. (2001). Renormalization of the voter model in equilibrium. *Annals of Probability*, 29(3):1262–1302.

# Index of Notation

**Numbers and spaces**

| | |
|---|---|
| $\mathbb{N}$ | natural numbers |
| $\mathbb{N}_0$ | $\mathbb{N} \cup \{0\}$ |
| $\mathbb{Z}$ | integers |
| $\mathbb{Q}$ | rational numbers |
| $\mathbb{R}$ | real numbers |
| $\mathbb{R}_+$ | nonnegative real numbers |
| $\mathbb{Z}^d$ | $d$-dimensional lattice |
| $\mathbb{R}^d$ | $d$-dimensional Euclidian space |
| $\Delta_I$ | the simplex in $\mathbb{R}^d$ |

**Sets**

| | |
|---|---|
| $\emptyset$ | the empty set |
| $A^c$ | the complement of $A$ |
| $\lvert A\rvert$ | the cardinality of $A$ |
| $\mathrm{ri}(A)$ | the relative interior of $A$ |
| $\partial^*(A)$ | the relative boundary of $A$ |
| $\mathrm{conv}(A)$ | the convex hull of $A$ |

**Special functions**

| | |
|---|---|
| $\Gamma$ | the $\Gamma$-function |
| $\Phi$ | the distribution function of a standard Gaussian random variable |

**Function spaces**

| | |
|---|---|
| $C(\mathbb{R})$ | the set of continuous functions on $\mathbb{R}$ |
| $C_b(\mathbb{R})$ | the set of continuous, bounded functions on $\mathbb{R}$ |
| $C^\psi$ | $\{f \in C(\mathbb{R}) :\ \lvert f\rvert \leq c \cdot \psi \text{ for some } c > 0\}$ |
| $B(\mathbb{R})$ | the set of bounded, measurable functions on $\mathbb{R}$ |
| $L^p$, $L^p(\Omega, \mathcal{F}, P)$ | $L^p$-spaces on the probability space $(\Omega, \mathcal{F}, P)$, $p \in [1, \infty]$ |

**Linear algebra**

| | |
|---|---|
| $\dim V$ | the dimension of the vector space $V$ |
| span $A$ | the vector space generated by $A$ |
| ker $M$ | the nullspace of the linear mapping $M$ |
| rg $M$ | the rank of the linear mapping $M$ |

**Probability and measures**

| | |
|---|---|
| $\mathcal{B}(E)$, $\mathcal{B}_E$ | the Borel $\sigma$-algebra on E |
| $\mathcal{M}_1(E)$ | the set of probability measures on E |
| $\mathcal{M}_1(\mathbb{R})$, $\mathcal{M}_1$ | the set of Borel probability measures on $\mathbb{R}$ |
| $\mathcal{M}_{1,c}(\mathbb{R})$, $\mathcal{M}_{1,c}$ | the set of Borel probability measures on $\mathbb{R}$ with compact support |
| $\mathcal{M}_c^+(\mathbb{R})$ | the set of Borel measures on $\mathbb{R}$ with compact support |
| $\mathcal{L}(X)$ | the law of $X$ |
| $X \sim \mu$ | $X$ has law $\mu$ |
| $\mathcal{L}(X\|\mathcal{F})$ | the conditional law of $X$ given $\mathcal{F}$ |
| supp $\mu$ | the support of $\mu$ |
| $\delta_x$ | Dirac measure on $x$ |
| $\mathbf{1}_A$ | indicator function of the set $A$ |
| $\text{unif}_A$ | uniform distribution on $A$ |
| $\mathcal{N}(a, \sigma^2)$ | normal distribution with mean $a$ and variance $\sigma^2$ |
| $\nu \ll \mu$, $\mu \gg \nu$ | $\nu$ is absolutely continuous with respect to $\mu$ |
| $\nu \approx \mu$ | $\nu$ and $\mu$ are equivalent |
| $\frac{d\nu}{d\mu}$ | the Radon-Nikodym derivative of $\nu$ with respect to $\mu$ |
| $F_\mu$ | the distribution function of $\mu$ |
| $F_\mu^{-1}$, $q_\mu$ | upper quantile function of $\mu$ |
| $H(\nu\|\mu)$ | the relative entropy of $\nu$ with respect to $\mu$ |

**Risk measures**

| | |
|---|---|
| $VaR_\lambda$ | value at risk at level $\lambda$ |
| $AVaR_\lambda$ | average value at risk at level $\lambda$ |
| $q_\lambda^-(\mu)$ | lower $\lambda$-quantile of $\mu$ |
| $q_\lambda^+(\mu)$ | upper $\lambda$-quantile of $\mu$ |